CSCL2011

computer-supported collaborative learning

July 4-8, 2011 Hong Kong, China

Connecting research to policy and practice

Conference Proceedings Volume I

Long papers

Editors: Hans Spada, Gerry Stahl, Naomi Miyake, Nancy Law

International Society of the Learning Sciences

Centre for Information Technology in Education
The University of Hong Kong

Connecting computer-supported collaborative learning to policy and practice

CSCL2011 CONFERENCE PROCEEDINGS
VOLUME I
LONG PAPERS

9th International Computer-Supported Collaborative Learning Conference

July 4-8, 2011, Hong Kong, China

Title:
Connecting Computer-Supported Collaborative Learning to Policy and Practice:
CSCL2011 Conference Proceedings Volume I — Long Papers

Editors:
Hans Spada, Gerry Stahl, Naomi Miyake, Nancy Law

Citation format:
Spada, H., Stahl, G., Miyake, N., Law, N. (Eds.) (2011) *Connecting Computer-Supported Collaborative Learning to Policy and Practice: CSCL2011 Conference Proceedings. Volume I — Long Papers*. International Society of the Learning Sciences.

ISBN:

ISLS
International Society of the Learning Sciences

http://www.isls.org/

Published by:
International Society of the Learning Sciences (ISLS)

Proceedings Printed by:
LuLu
http://www.lulu.com

Proceedings Distributed by:
LuLu and Amazon

Book Cover Design:
Debbie Pang, Multimedia Web Designer
Centre for Information Technology in Education (CITE), The University of Hong Kong
http://www.cite.hku.hk/

Conference Organizers and Supporting Organization

Organizers

- The International Society of the Learning Sciences
- Centre for Information Technology in Education, The University of Hong Kong

Co-organizers

- Beijing Normal University
- East China Normal University
- South China Normal University

Supporting Organization

- Education Bureau, the Hong Kong Special Administrative Region

Conference Sponsors

Platinum Sponsor

- Strategic Research Theme (SRT) on the Sciences of Learning, The University of Hong Kong

Gold Sponsors

- Hong Kong Applied Science and Technology Research Institute
- Sunny Interactive

Other Sponsor

- Bank of East Asia

Steering Committee

Conference Chairs
- Nancy Law, The University of Hong Kong, China
- Kai Ming Cheng, The University of Hong Kong, China

Program Chairs
- Hans Spada, Freiburg University, Germany
- Gerry Stahl, Drexel University, USA
- Naomi Miyake, University of Tokyo, Japan

Pre-conference Workshops and Tutorials Chairs
- Daniel Suthers, University of Hawai'i, USA
- Iris Tabak, Ben Gurion University of the Negev, Israel
- Chee Kit Looi, Nanyang Technological University, Singapore

Post-conference Events and Tutorials Chairs
- Ronghuai Huang, Beijing Normal University, China
- Shengquan Yu, Beijing Normal University, China
- Zhiting Zhu, East China Normal University, China
- Youqun Ren, East China Normal University, China
- Fuyin Xu, South China Normal University, China
- Hongguang Peng, Guangdong Center of Educational Technology, China
- Baohui Zhang, National Institute of Education, Singapore
- Siu Cheung Kong, The Hong Kong Institute of Education
- Nancy Law, The University of Hong Kong

Doctoral Consortium Chairs
- Kristine Lund, University of Lyon, France
- Christopher Hoadley, New York University, USA
- Carol Chan, The University of Hong Kong

Early Career Workshop Chairs
- Frank Fischer, University of Munich, Germany
- Cindy Hmelo-Silver, Rutgers University, USA
- Tak Wai Chan, National Central University, Chinese Taipei

Interactive Events, Demonstrations and CSCL in Practice Showcases Chairs
- Therese Laferriere, University of Laval, Canada
- Katerine Bielaczyc, Nanyang Technological University, Singapore
- Sadhana Puntambekar, University of Wisconsin-Madison, USA
- Nancy Songer, University of Michigan, USA
- Sanna Javela, University of Oulu, Finland

Community Memory Chair
- Ricki Goldman, New York University, USA
- Diane Stormont, University of Hong Kong
- Kevin Lau, University of Hong Kong

International Relations Chairs
- Peter Reimann, University of Sydney, Australia
- Tak Wai Chan, National Central University, Chinese Taipei
- Paivi Hakkinen, University of Jyvaskyla, Finland

Advisors
- Ying Chan, The University of Hong Kong
- Paul Kirschner, Open University, Netherlands
- Claire O'Malley, University of Nottingham, UK
- Janet Kolodner, Georgia Institute of Technology, USA
- Daniel Suthers, University of Hawai'i, USA

Local Organizing Committee (The University of Hong Kong)

Local Organization Chair
- Prof. Nancy Law

Paper Submission and Review Process Technological System Chair
- Mr. Murphy Wong
- Mr. Simon Leung

Graphic Designer and Website Chair
- Ms. Debbie Pang

Proceedings Editing Chairs
- Dr. Man Wai Lee
- Dr. Tamara Savelyeva

Technology Support and Registration System Chair
- Mr. Murphy Wong
- Mr. Geoffrey R. Dengate

Mainland China Post-conference Coordinators
- Dr. Jingyan Lu
- Dr. Ming Lai

Policy Forum Coordinator
- Dr. Jae Park

Interactive Events Coordinators
- Dr. Yeung Lee
- Mr. Johnny Yuen

Pre-conference Events Coordinator
- Mr. Ron Tsoi

Posters Coordinator
- Mr. Ron Tsoi

Community Memory Chairs
- Dr. Bob Fox
- Dr. Felix Siu
- Mr. Simon Leung

Program Booklet Editing Chair
- Dr. Allan Yuen

Registration and Attendee Materials Chair
- Ms. Bella Chow

Running Conference Coordinator
- Ms. Candy Leung

Social Events Chairs
- Ms. Candy Leung
- Dr. Jae Park

Students Volunteer Chairs
- Dr. Alvin Kwan
- Mr. Peter Warning

Treasurer
- Ms. Candy Leung

Mainland China Post-conference

Local Organizing Committee (Guangzhou)

Chairs

- Fuyin Xu
  South China Normal University
- Jun Xu
  South China Normal University
- Hongguang Peng
  Guangdong Center of Educational Technology

Members

- Xionghui Chen
  South China Normal University
- Jianli Jiao
  South China Normal University
- Junfen Lin
  Guangdong Center of Educational Technology
- Jianhua Zhao
  South China Normal University
- Qianwei Zhang
  South China Normal University
- Su Mu
  South China Normal University

Local Organizing Committee (Shanghai)

Chairs

- Zhiting Zhu
  East China Normal University
- Youqun Ren
  East China Normal University
- Pengkai Pan
  Saybot Information Technology (Shanghai) Co., Ltd.

Members

- Xinning Pei
  East China Normal University
- Xianze Cai
  Saybot Information Technology (Shanghai) Co., Ltd.
- Hanbing Yan
  East China Normal University
- Jian Zhao
  East China Normal University
- Xiaoqing Gu
  East China Normal University
- Bin He
  East China Normal University

Local Organizing Committee (Beijing)

Chairs

- Ronghuai Huang
  Beijing Normal University
- Shengquan Yu
  Beijing Normal University

Members

- Mang Li
  Beijing Normal University
- Yanyan Li
  Beijing Normal University
- Liang Yu
  Beijing Normal University
- Yanqin Zheng
  Beijing Normal University

REVIEWERS OF CSCL2011 CONFERENCE

The conference chairs and program chairs sincerely thank the following 293 scientists who reviewed proposals for the CSCL2011 conference.

Abrahamson, Dor, University of California, Berkeley
Abramovich, Sam, University of Pittsburgh
Acholonu, Ugochi, Stanford University
Albury, Steven, University of Oxford
Aleahmad, Turadg, Carnegie Mellon University
Alterman, Richard, Brandeis University
Anjewierden, Anjo, University of Twente
Arena, Dylan, Stanford University
Arnseth, Hans Christian, Department of Education, University of Oslo
Asensio, Juan I., University of Valladolid
Avouris, Nikolaos, University of Patras
Bader-Natal, Ari, Grockit, Inc.
Bagley, Elizabeth, University of Wisconsin-Madison
Bairral, Marcelo, Federal Rural University of Rio de Janeiro (UFRRJ)
Baker, Michael, CNRS - Telecom ParisTech
Bauters, Merja, Helsinki Metropolia University of Applied Sciences
Bennerstedt, Ulrika, University of Gothenburg
Bermundo, Cesar, Ateneo de naga University
Blavier, Adelaide, University of Liège
Bodemer, Daniel, University of Tubingen
Bouyias, Yannis, Aristotle Univesity of Thessaloniki
Bratitsis, Tharrenos, University of Western Macedonia, Greece
Brennan, Karen, MIT Media Lab
Bruckman, Amy, Georgia Tech
Buder, Juergen, KMRC Tuebingen
Bures, Eva, Bishop's University
Cakir, Murat, Drexel University
Carell, Angela, University of Bochum
Cavallo, Alice, Harvard
Chan, Carol, University of Hong Kong
Chan, Margaret, Columbia University
Chan, Tak-Wai, National Central University
Chang, Ben, National Chiayi University
Chapman, Robbin, Massachusetts Institute of Technology
Charles, Elizabeth, Dawson College
Chau, Clement, Tufts University
Chen, Ching-Huei, National Changhua University of Education
Chen, Fei-Ching, National Central University, NCU
Chen, Wenli, National Institute of Education
Cherniavsky, John, National Science Foundation
Ching, Yu-Hui, Boise State Univerity
Churchill, Daniel, University of Hong Kong
Clarke-Midura, Jody, Harvard Graduate School of Education
Clegg, Tamara, Georgia Institute of Technology
Code, Jillianne, Harvard University
Condamines, Thierry, Laboratoire MIS
Cress, Ulrike, Knowledge Media Research Center
Daily, Shaundra, g8four
Damsa, Crina, University of Oslo
Dascalu, Mihai, University Politehnica of Bucharest
Dasgupta, Chandan, University of Illinois at Chicago
de Jong, Frank, Stoas University of Applied Sciences and teacherr Education
de Leng, Bas, Maastricht University
Deiglmayr, Anne, University of Freiburg
Demetriadis, Stavros, Aristotle University of Thessaloniki, Greece
Demmans Epp, Carrie, University of Saskatchewan
DeVane, Ben, University of Florida
Di Blas, Nicoletta, Politecnico di Milano
Dillenbourg, Pierre, EPFL
Dimitriadis, Yannis, University of Valladolid
Dirckinck-Holmfeld, Lone, Aalborg University
DiSalvo, Betsy, Georgia Insitute of Technology
Diziol, Dejana, University of Freiburg
Dowell, John, University College London
Dugstad Wake, Jo, University of Bergen, UiB
Duncan, Sean, Miami University
Dyke, Gregory, Ecole Nationale Supérieure des Mines de Saint-Etienne
Engelmann, Tanja, Knowledge Media Research Center
Erlandson, Benjamin, California State University, Monterey Bay
Ertl, Bernhard, Universität der Bundeswehr München
Evans, Michael, VIrginia Tech
Feldmann, Birgit, University of Hagen
Fesakis, Georgios, University of Aegean
Fields, Deborah, University of California, Los Angeles
Filsecker, Michael, Duisburg-Essen University
Fischer, Frank, University of Munich
Fischer, Gerhard, University of Colorado, Boulder
Forte, Andrea, Drexel University
Fox, Bob, University of Hong Kong
Fujita, Nobuko, OISE/UT
Fuks, Hugo, Catholic University of Rio de Janeiro
Gegenfurtner, Andreas, University of Turku

George, Sebastien, INSA de Lyon
Goggins, Sean, Drexel University College of Information Science & Technology
Gogoulou, Agoritsa, Department of Informatics & Telecommunications, University of Athens
Gouli, Evangelia, University of the Aegean
Grant, Jamillah, Northcentral University
Gressick, Julia, University of Wisconsin- Madison
Guribye, Frode, University of Bergen
Häkkinnen, Päivi, University of Jyväskylä, Finnish Institute for Educational Research
Haake, Joerg, FernUniversitaet in Hagen
Hackbarth, Alan, University of Wisconsin-Madison
Halverson, Erica, University Of Wisconsin - Madison
Hayama, Tessai, Japan Advanced Institute of Sience and Technology
Heo, Gyeong Mi, CEFRIO
Hernandez, Juan Carlos, Open University Colombia UNAD
Herodotou, Christothea, Cyprus University of Technology
Herrmann, Thomas, University of Bochum, Germany
Hilbert, Tatjana, University of Mainz
Hmelo-Silver, Cindy, Rutgers University
Hoidn, Sabine, Harvard University, School of Education
Hong, Huang-Yao, National Chengchi University
Hong, Kian Sam, Universiti Malaysia Sarawak
Honwad, Sameer, Rutgers University
Hoppe, H. Ulrich, University of Duisburg-Essen
Horn, Michael, Northwestern University Learning Sciences
Horney, Mark, University of Oregon
Hsu, Yu-Chang, Boise State University
Hubscher, Roland, Bentley University
Hui, Diane, The University of Hong Kong
Ioannou, Andri, Cyprus University Of Technology
Iordanou, Kalypso, Neapolis University
Jacobson, Michael, The Univerisity of Sydney
Jahnke, Isa, Dortmund University of Technology
Janssen, Jeroen, Utrecht University
Jeong, Heisawn, Hallym University
Joiner, Richard, University of Bath
Juang, Yih-Ruey, Jinwen University of Science and Technology
Kafai, Yasmin, ucla education
Kali, Yael, University of Haifa
Kang, Haijun, Jackson State University
Kaptelinin, Victor, University of Bergen, Norway
Kapur, Manu, National Institute of Education
Karakostas, Anastasios, Aristotle University of Thessaloniki
Khan, Azam, Autodesk, Inc.
Khine, Myint Swe, Emirates College for Advanced Education
Kim, Beaumie, National Institute of Education
Kim, Kibum, Virginia Tech
Kim, Mi Song, Nanyang Technological University
Kimmerle, Joachim, University of Tuebingen, Department of Applied Cognitive Psychology and Media Psychology
Kirschner, Paul, Open University of the Netherlands
Knipfer, Kristin, University of Tuebingen
Koh, Myung, Riverside City College
Koops, Willem, Máxima Medical Center
Kopp, Birgitta, University of Munich
Koumpis, Adamantios, ALTEC
Kraemer, Nicole, University Duisburg-Essen
Krange, Ingeborg, InterMedia, University of Oslo
Krumm, Andrew, University of Michigan
Kumar, Swapna, University of Florida
Kwon, Samuel, Concordia University Chicago
Kyza, Eleni, Cyprus University of Technology
Lütticke, Rainer, Bochum University of Applied Sciences
Laferriere, Therese, Laval University
Lafifi, Yacine, LAIG Laboratory, Guelma University, Algeria
Lai, Ming, The University of Hong Kong
Lakkala, Minna, University of Helsinki
Lambropoulos, Niki, Intelligenesis
Lan, Yu-Ju, National Taiwan Normal University
Lansiquot, Reneta, New York City College of Technology
Larusson, Johann, Brandeis University
Lazakidou, Georgia, University of Piraeus
Levy, Dalit, HIT
Li, Ken W., Hong Kong Institute of Vocational Education (Tsing Yi)
Li, wenjuan, University of Illinois at Chicago
Li, Xiao, Southwest China University
Liao, Chang-Yen, Graduate Institute of Network Learning Technology, National Central University
Lim, Jung, Northwestern State University
Lindgren, Robb, University of Central Florida
Lindwall, Oskar, Göteborg University
Lingnau, Andreas, University of Strathclyde
Lipponen, Lasse, University of Helsinki
Liu, Lei, University of Pennsylvania
Lonchamp, Jacques, LORIA - University of Nancy
Lonn, Steven, University of Michigan
Looi, Chee-Kit, Nanyang Technological University
Lorek, Edward, Naval Air Systems Command 4.6
Lu, Jingyan, The University of Hong Kong
Luckin, Rosemary, The London Knowledge Lab
Ludvigsen, Sten, InterMedia
Lund, Andreas, University of Oslo
Lund, Kristine, ICAR
Lyons, Leilah, University of Illinois at Chicago
Mäkitalo-Siegl, Kati, University of Jyväskylä

Mahardale, Jay, Admiralty Primary School, Ministry of Education
Makara, Kara, University of Michigan
Manca, Stefania, Institute for Educational Technology, Italian National Research Council
Markauskaite, Lina, The University of Sydney
Martínez-Monés, Alejandra, University of Valladolid
Mavrikis, Manolis, Institute of Education
McLaren, Bruce, Carnegie Mellon University
Medina, Richard, Laboratory for Interactive Learning Technologies
Mercier, Emma, Durham University
Mochizuki, Toshio, Senshu University
Molinari, Gaelle, Distance Learning University Switzerland
Mor, Yishay, London Knowledge Lab
Moraveji, Neema, Stanford University
Morch, Anders, InterMedia. University of Oslo
Muehlpfordt, Martin, FernUniversitaet Hagen
Mwanza-Simwami, Daisy, The Open University
Najafi, Hedieh, Ontario Institute for Studies in Education
Noroozi, Omid, Wageningen University
Notari, Michele, School of Teacher Education PHBern
Nwaigwe, Adaeze, American University of Nigeria
Oehl, Michael, Leuphana University of Lueneburg
Ogata, Hiroaki, University of Tokushima
Oshima, Jun, Shizuoka University
Oshima, Ritsuko, Shizuoka University
Palmquist, Sasha, Institute for Learning Innovation
Pan, Rui, Purdue University
Papanikolaou, Kyparisia, School of Pedagogical and Techological Education, ASPETE
Pea, Roy, Stanford University
Pemberton, Lyn, University of Brighton
Penuel, William, SRI International
Pfister, Hans-Rüdiger, University of Lüneburg
pierroux, palmyre, university of oslo
Pifarré, Manoli, Universitat de Lleida
Pinkwart, Niels, Clausthal University of Technology
Price, Sara, Institute of Education
Rebedea, Traian, Politehnica University of Bucharest
Reedy, Gabriel, King's College London
Reever, Richard, Queen's University
Reimann, Peter, University of Sydney
Renninger, K. Ann, Swarthmore College
Richter, Christoph, Christian-Albrechts-Universität zu Kiel
Rick, Jochen, Saarland University
Roberts, Tim, Central Queensland University
Rogers, Jim, Utah State University
Roll, Ido, University of British Columbia
Romero, Margarida, Université of Toulouse, CLLE-LTC (UMR 5263)
Roschelle, Jeremy, SRI International
Rose, Carolyn, Carnegie Mellon University
Rummel, Nikol, Ruhr-Universittät Bochum
Rusman, Ellen, Open University of the Netherlands
Salden, Ron, Madeira Interactive Technologies Institute
Saleh, Asmalina, Indiana University Bloomington
Schmidt, Matthew, University of Missouri
Schroeder, Ulrik, RWTH Aachen University
Schwendimann, Beat, UC Berkeley
Sethi, Ricky, UCLA
Shaoming, Chai, South china Normal University
Shirouzu, Hajime, Chukyo University
Sie, Rory, Open Universiteit in the Netherlands
Simon, Jean, Institut Universitaire de Formation des Maitres
Siyahhan, Sinem, Indiana University
Slof, Bert, University of Groningen
Smith, Garrett, University of Wisconsin - Madison
So, Hyo-Jeong, National Institute of Education, Nanyang Technological University
Stegmann, Karsten, University of Munich
Steinkuehler, C, University of Wisconsin-Madison
Stoyanov, Slavi, Open University of the Netherlangs (OUNL)
Strijbos, Jan-Willem, Leiden University
Subramaniam, Nantha Kumar, Open University Malaysia
Suthers, Dan, University of Hawaii
Svihla, Vanessa, The University of Texas at Austin
Sylvan, Elisabeth, MIT Media Lab
Tabak, Iris, Ben Gurion University
Tartaro, Andrea, Union College
Tchounikine, Pierre, LIG
Teasley, Stephanie, University of Michigan
Teplovs, Christopher, OISE/UT
Thomas, Jakita, Spelman College
Trausan-Matu, Stefan, Politehnica University of Bucharest
Trninic, Dragan, University of California, Berkeley
Tsiatsos, Thrasyvoulos, Aristotle University of Thessaloniki
Tzagarakis, Manolis, Research Academic Computer Technology Institute
Underwood, Jody, Pragmatic Solutions, Inc.
Underwood, Joshua, London Knowledge Lab
van Aalst, Jan, The University of Hong Kong
Van Amelsvoort, Marije, Tilburg University
van de Sande, Carla, Arizona State University
van Leeuwen, Anouschka, Utrecht University
van 't Hooft, Mark, Kent State University
Vanover, Charles, University of South Florida Saint Petersburg
Vatrapu, Ravi, Copenhagen Business School
Verdines, Patricia, Tecnologico de Monterrey
Vignollet, Laurence, Université de Savoie

Vu Minh, Chieu, University of Michigan, School of Education
Walker, Erin, Carnegie Mellon University (CMU)
Wang, Xinghua, University of Munich
Wardrip, Peter, University of Pittsburgh
Warner, Norman, Naval Air Systems Command
Wecker, Christof, University of Munich, Department of Psychology
Wee, Juan Dee, National Institute of Education, Nanyang Technological University
Wessner, Martin, Fraunhofer IESE
White, Tobin, UC Davis
Wilkerson-Jerde, Michelle, Northwestern University
Winter, Marcus, University of Brighton
Wise, Alyssa, Simon Fraser University
Wong, Lung Hsiang, National Institute of Education
Woo, David, University of Hong Kong
Wopereis, Iwan, Centre for Learning Sciences and Technologies - Open University of the Netherlands
Xhafa, Fatos, Technical University of Catalonia
Yacef, Kalina, School of Information Technologies
Yeo, Jennifer, National Institute of Education
Yoo, Jaebong, USC/ISI
Yoon, Susan, University of Pennsylvania
Yu, Fu-Yun, National Cheng Kung University
Yuen, Johnny K.L., University of Hong Kong
Zahn, Carmen, Knowledge Media Research Centre
Zhang, Jianwei, University at Albany, SUNY
Zhou, Nan, Drexel University
Zourou, Katerina, University of Luxembourg

Introduction to the Proceedings of CSCL 2011

Gerry Stahl, Hans Spada, Naomi Miyake, Nancy Law

The Scientific Field of CSCL

Computer-Supported Collaborative Learning (CSCL) is a multidisciplinary research field inspired by the power of collaborative learning and by the promise of computer technologies to support collaborative learning. It draws on and explores constructivist and socio-cultural theories, which view learning as a social, interpersonal, meaning-making process that takes place largely through interaction among people and within communities. It also designs, adopts and refines technologies that mediate communication among learners and that help to guide their inquiry or structure their work.

As a research field, CSCL builds on conceptual frameworks and analytic approaches of many academic fields, including education, psychology, communication, computer science and social science. It applies a variety of quantitative and qualitative research methods, often combining them to develop richer understandings of complex phenomena. Likewise, it may involve both laboratory and classroom studies, formal and informal learning settings, different temporal scales and the study of a wide range of influential factors.

Policies and Practices for CSCL

While the CSCL conference series has centered on research studies, the field has always been strongly oriented toward practical concerns of educational practice and associated educational policy. CSCL research frequently involves teachers in school classrooms and seeks to influence or implement governmental education policies.

The CSCL 2011 conference theme, "*Connecting computer-supported collaborative learning to policy and practice*," builds on previous CSCL conferences to examine whether and how CSCL practices can bring deep changes to formal and informal educational practices at all levels, and contribute to educational improvement at a system level by informing education policy. This theme is addressed by keynote talks, symposia, trips to schools, and other events at the conference and the post-conference. It is hoped that this conference theme will contribute to bringing greater recognition to the fields of CSCL and the Learning Sciences by drawing the attention of a wider public, including policy makers and the professional educational community to their research and development contributions.

One important feature of this year's conference is the inclusion of three parallel tracks of interactive events, demonstrations and CSCL-in-practice showcases, which serve as the foci for attracting practitioners to the conference. Included in these practitioner-oriented events are presentations from several prominent school-university partnership projects that are themselves good exemplars of the conference theme in action. The conference has the support of policy makers in Hong Kong to sponsor teacher participation at the conference; the Education Bureau of the HKSAR Bureau is a supporting organization for this conference. The practitioner tracks are also made possible through the merger of other conferences into this year's CSCL conference. This year, the annual Knowledge Building Summer Institute, which has usually been held in Toronto, Canada, has been integrated into the CSCL conference in Hong Kong and Guangzhou.

To take advantage of CSCL 2011 being held in Hong Kong, CITE collaborated with East China Normal University, South China University and Beijing Normal University to co-organize a series of CSCL 2011 post-conference events in Shanghai, Guangzhou and Beijing respectively on July 11-15. It is the first time that there are such major post-conference events for the CSCL conference and we hope this will provide more opportunities for academic exchange and collaboration between CSCL and learning sciences researchers in Greater China and their global counterparts.

The CSCL Community and Conference

Since 1995, the CSCL conference has provided a stimulating and friendly venue for people interested in the multi-disciplinary issues of computer-supported collaborative learning to meet in a relaxed atmosphere with a variety of formal and informal events. Structured activities and social occasions promote interpersonal relations and knowledge building. The conference's human size and structure facilitate getting to know international colleagues and discussing cutting-edge ideas in educational practice, technology design, CSCL theory and diverse research approaches.

The bi-annual conferences have been instrumental in developing the field of CSCL and in building the research community around it. The conferences took place in Bloomington, USA (1995), Toronto, Canada

(1997), Stanford, USA (1999), Maastricht, Netherlands (2001), Boulder, USA (2002), Bergen, Norway (2003), Taipei, Taiwan (2005), New Brunswick, USA (2007) and Rhodes, Greece (2009).

Further efforts to build the CSCL field include the founding of the International Society of the Learning Sciences (ISLS) by the CSCL community and the Learning Sciences research community. ISLS now provides an institutional framework for running the CSCL and ICLS conferences in alternating years and for publishing the *International Journal of Computer-Supported Collaborative Learning* (*ijCSCL*) and the *Journal of the Learning Sciences* (*JLS*). In the early days of CSCL research, there was no publication venue specifically oriented to the field and it was hard to locate publications in the field. Now, in addition to the CSCL journal, there is also a CSCL book series sponsored by ISLS and published by Springer. Furthermore, papers from the CSCL and ICLS conferences are available in the ACM Digital Library and both *ijCSCL* and *JLS* are abstracted in the major indexing services, where they are highly ranked.

Toward a Global CSCL

The first CSCL conference was a relatively simple event, held in the middle of the United States. Over the years, the conference expanded to include a variety of sessions to meet the needs of a growing research community. It now features long papers presented lecture style, posters presented interactively and short papers presented in a hybrid style, to accommodate research findings ranging from early work to more mature reports. There are also tutorials for newcomers and workshops for special hot topics. For doctoral students and new faculty, there is a doctoral consortium and an early career workshop. There are also opportunities for software demos and other interactive events. And of course there are receptions and other social events to give extra times for people to get to know each other.

Although the CSCL community always had a strong base in Western Europe—partially associated with the AI and Education community—the first official CSCL conferences were held in North America. In 2001, a Euro-CSCL conference was organized in the Netherlands, attracting mainly European researchers. In 2002, the conference in the US achieved a good balance of European and American researchers; it initiated a policy of rotating the conferences to Europe (in 2003 and 2009), Asia (in 2005 and 2011) and North America (in 2007 and 2013). The conference in Taipei (2005) succeeded in achieving a good balance of paper authors, program committee members and conference participants from Western Europe, North America and the Asia-Pacific region.

Internationalization has always been a goal of the CSCL community. An analysis of trends during the first decade of the conferences documented strong progress in that direction (Kienle & Wessner, 2006). Analysis of authors included in the CSCL 2011 main conference shows approximately equal participation from Western Europe, North America and Asia-Pacific. Another important trend is an increase in the number of international collaborations in research and in the co-authorship of papers reporting on that research. Such collaboration is necessary for the spread of expertise and deep understanding of innovative ideas, methods and tools. This year's post-conference activities are an additional opportunity to promote exchange with researchers, practitioners and policy makers in Mainland China, an important area in which CSCL approaches seem to be spreading rapidly.

Of course, there are still major regions of the world under-represented in the CSCL community, such as the Middle East, Eastern Europe, Central Asia, South America and Africa. To some extent this may be due to limited traditions of collaborative learning or relatively low levels of computerization in schools in those areas. It may also be due to limitations in resources for traveling to international conferences or in awareness of the field. We have seen that strong involvement in CSCL research generally requires policy initiatives backed up with funding commitments. The European Union Network of Excellence funding programs like Kaleidoscope and Stellar have made a significant difference. NSF support for educational research has helped in the USA as well. Case studies elsewhere underline this factor (Chan, 2011; Looi et al., 2011).

A Delphi survey of researchers and stakeholders in technology-enhanced learning recently ranked CSCL as the second most important core research area for the next decade—just behind "connection between informal and formal learning" and ahead of nine other areas, like "personalized learning" (Kaendler et al., these Proceedings, Vol. II). We hope this recognition will spread around the world. In order to address the challenges facing CSCL in the coming years—not least of which are those related to practice and policy—we need the combined efforts of a global collaborative effort. Such an effort would bring together the unique perspectives of many labs and diverse educational cultures, acknowledging and strengthening their individual perspectives while incorporating them into a global synthesis.

Volume I: CSCL 2011 Long Papers

Volume I of the Proceedings includes the papers that were accepted through peer review for presentation as long papers. These papers were submitted in November 2010 and were reviewed by three anonymous reviewers. A member of the Program Committee then summarized the three reviews and wrote a recommendation. The three Program Committee Co-Chairs considered the reviews and recommendations—and in many cases read the papers. Based on this, they agreed on a list of 72 submissions to accept as long papers, grouping them into 18 sets of 4 thematically related papers that could be presented in the scheduled long-paper sessions. Out of 188 submissions of long papers, 72 (38%) were accepted as long papers, 45 (24%) as short papers, 48 (26%) as posters and 23 (12%) were rejected.

Volume II: CSCL 2011 Short Papers and Posters

Volume II of the Proceedings includes the papers that were accepted through peer review for presentation as short papers or posters. Submissions for short papers or posters went through exactly the same review process as long papers. Out of 52 submissions of short papers, 17 (33%) were accepted as short papers, 26 (50%) as posters and 9 (17%) were rejected. Out of 38 submissions of posters, 26 (68%) were accepted as posters and 12 (32%) were rejected. Short papers were grouped into sets of 6 thematically related papers. Authors of short papers give very brief presentations and then conduct round-table discussions of their papers with interested audience members. Posters were assigned to two poster sessions; authors of posters hang large-scale posters on walls and then discuss them with interested audience members.

Volume III: CSCL 2011 Pre-Conference, Keynotes, Symposia and Post-Conference

Volume III of the Proceedings includes summaries of other events at CSCL 2011.

The pre-conference events include workshops, tutorials, a Doctoral Consortium and an Early Career Workshop. There are three tutorials on tool support for analysis, social network analysis and the WISE environment. There are four workshops on orchestrating CSCL in the classroom, connecting levels of learning and synthesizing three approaches to CSCL design.

The highlights of this year's conference include keynote talks by prominent speakers: Dr. Gwang-Jo Kim, Director of UNESCO Regional Bureau for Education in Asia-Pacific; Dr. Ed H. Chi, Research Scientist, Google Research; Prof. Erik Duval, Professor of Computer Science, Katholieke Universiteit Leuven, Belgium; and Prof. Roy Pea, Stanford University Professor of the Learning Sciences and Director of the Stanford Center for Innovations in Learning. The keynotes cover the full range of issues from researcher, policy-maker and practice perspectives.

Out of 17 proposals for symposia, 14 (82%) were accepted and 3 (18%) were rejected. This high acceptance rate of symposia is a result of the fact that most of them had been carefully filtered by large groups of organizers. The submissions were generally of exceptional quality and represented important and timely themes that are of current high relevance to the field. They often reflect important centers of CSCL research in different regions of the world or international collaborations. In order to avoid having these symposia draw audiences away from long and short paper sessions, the symposia were mostly scheduled against each other.

The practitioner-oriented sessions take place in parallel with the paper and symposium sessions of the main conference. They include a wide variety of presentations and events that are designed for classroom teachers and others particularly interested in the applications of CSCL research and their use in the classroom. This strand of activities showcases design research in CSCL involving field-based educators and/or strong university-school partnerships. These events are of interest and benefit to teachers and other practitioners, as well as researchers and educators interested in models and exemplars of research and practice interaction and partnership. They are listed in the conference Program.

The post-conference consists of a series of conference activities to be held in Shanghai, Guangzhou and Beijing in China. It builds on the conference theme of connecting CSCL research to education policy and practice. It draws on national and global exemplars of synergistic advances in CSCL and learning sciences research and educational policy and practice to explore the current state and the way forward for education developments in China. This series of post-conference activities brings together researchers, practitioners and policy-makers within China and internationally to identify ways to better leverage the potentials that research on learning and learning technologies bring to educational change and improvement.

Hong Kong University Centenary

The CSCL 2011 conference coincides with a major local milestone as well as an advance of the CSCL community. A century ago, in 1911, the University of Hong Kong was incorporated by Ordinance. A group of visionaries founded the first university in Hong Kong, from which generations of leaders across the region would come forth. The University of Hong Kong was to be important for China and for the world. In celebrating the first centenary, HKU upholds its commitment to Knowledge, Heritage and Service. The Centre for Information Technology in Education (CITE) of the Faculty of Education is proud to be hosting the CSCL 2011 main conference and co-organizing the CSCL 2011 post-conferences in three Mainland Chinese cities as part of the HKU Centenary celebration events.

References

Chan, C. (2011). CSCL theory-research-practice synergy: The Hong Kong experience of implementing knowledge building in classrooms. *International Journal of Computer-Supported Collaborative Learning. 6*(2). Doi: http://dx.doi.org/10.1007/s11412-011-9121-0.

Kienle, A., & Wessner, M. (2006). The CSCL community in its first decade: Development, continuity, connectivity. *International Journal of Computer-Supported Collaborative Learning. 1*(1), 9-33. Doi: http://dx.doi.org/10.1007/s11412-006-6843-5.

Looi, C.-K., So, H.-j., Toh, Y., & Chen, W. (2011). CSCL in classrooms: The Singapore experience of synergizing policy, practice and research. *International Journal of Computer-Supported Collaborative Learning. 6*(1), 9-38. Doi: http://dx.doi.org/10.1007/s11412-010-9102-8.

Table of Contents

VOLUME I PART 2: LEARNING ACTIVITIES AND TOOLS

VOLUME I PART 3: LEARNING INTERACTIONS

VOLUME I PART 4: TECHNOLOGY-ENHANCED INTERACTIONS AND ANALYSIS

VOLUME I PART 5: ISSUES IN CSCL

(END OF VOLUME I)

VOLUME II: SHORT PAPERS & POSTERS

VOULME II PART 1: KNOWLEDGE BUILDING

VOLUME II PART 2: LEARNING ACTIVITIES AND TOOLS

VOLUME II PART 3: LEARNING INTERACTIONS

VOLUME II PART 4: TECHNOLOGY-ENHANCED INTERACTIONS & ANALYSIS

VOLUME II PART 5: ISSUES IN CSCL

VOLUME II POSTERS

(END OF VOLUME II)

VOLUME III: KEYNOTES, SYMPOSIA, PRACTITIONER-ORIENTED EVENTS, PRE-CONFERENCE AND POST-CONFERENCE

VOLUME III PART 2: SYMPOSIA 1013

(END OF VOLUME III)

Volume I

Long papers

Construction of Health Knowledge in an Alternative Medical Community of Practice: Hermeneutic Analysis of a Web Forum

Joachim Kimmerle, Kristina Abels, Katharina Becher, Anna Beckers, Annette Haussmann, Ansgar Thiel,
University of Tuebingen, 72072 Tuebingen, Germany
Email: j.kimmerle@iwm-kmrc.de, kristinaabels@gmx.de, katharina.becher@student.uni-tuebingen.de, anna.beckers@student.uni-tuebingen.de, annette.haussmann@web.de, ansgar.thiel@uni-tuebingen.de
Ulrike Cress, Knowledge Media Research Center, 72072 Tuebingen, Germany, u.cress@iwm-kmrc.de

Abstract: This paper presents a hermeneutic analysis of a web forum in which supporters of an alternative medical approach, the so-called *Urkost* ("primordial food") movement, are gathered as a community of practice, and in which they exchange information and opinions in order to pass on and develop relevant knowledge. In our analysis of knowledge construction, we were able to identify four aspects. 1) Imparting a certain view of the world – knowledge creation provided an ideological guidance system. 2) Role structure within the community – the entire communication was centered on one individual expert who played the role of a kind of guru. 3) Persuasion and reinforcement of followers – the impartment of knowledge and persuasion were primarily based on an emotional type of communication. 4) Separation from "dissidents" – the attitude towards people of a "different faith" and towards standard medicine was characterized by strong hostility and rejection.

Introduction

Medical progress in Western civilizations has brought various achievements in the past decades. But the specialization and differentiation that accompanied these achievements is perceived by many patients and people who are seeking help as unpleasant and inappropriate, because often they feel they are being reduced to their symptoms, and have the impression of being left alone with their worries. Not surprisingly, many people prefer alternative medical approaches, which appear to take a more holistic view of human beings, and present solutions to health problems in simple language. This trend may be considered in the context of the development of cultic milieus in modern societies in general (Campbell, 1972) and medically oriented cultic milieus in particular (Salmon, 1984). New media make it very easy to catch up on what is known about diseases and treatments that represent alternative medical methods. For this purpose, a variety of different alternative medical communities have come together on the Internet, where they exchange their respective information. This paper explores how such communities create and develop their health knowledge. In the following section, one (quite radical) alternative medical community, the so-called *Urkost* movement, will be introduced. This community was selected for analysis because of three reasons: 1) This extreme movement represents a typical example of a cultic milieu, bearing in mind that its basic assumptions are highly controversial from a scientific point of view, but – nevertheless – it has a significant number of compliant followers. 2) The *Urkost* community is interesting as it propagates, on the one hand, an "archaic" lifestyle, but uses, at the same time, the Internet, which is really a typical post-modern form of communication. 3) This group has some characteristic features of a community of practice (see below), and the interesting question here is whether their development of knowledge occurs in a similar way as in a scientific community. After the introduction of the *Urkost* community we will describe the method that was applied here. In the main part, we will then present the four modalities of knowledge creation in the *Urkost* community, as they emerged from the data in our analysis: imparting a certain world view; role structure within the community; persuasion of newcomers and skeptics; and separation from standard medicine and people of some "different faith". In the conclusion, our findings will be summarized and discussed.

The *Urkost* Movement As an Alternative Medical Community of Practice

The term *alternative medicine* (Bratman, 1997) refers to all those procedures and methods of treatment in the field of medicine which are not (or only to some limited extent) accepted by conventional medicine, because their claimed effects cannot be substantiated using scientific methods. Alternative medical communities are characterized by a shared interest in a certain topic. The members of such communities look into the subject of their shared interest and deal with it; and in this process they develop corresponding competencies. In this context, knowledge development is supposed to take place at both an individual and a collective level (Kimmerle, Cress, & Held, 2010). The *Urkost* movement (the German term *Urkost* may best be translated as "primordial food") was established by Franz Konz (*1926), who was at one time well known in Germany as an author of handbooks on how to save tax. In the first instance, *Urkost* is a nutrition concept that says that all diseases are a consequence of some lifestyle that is not in harmony with nature. Nutrition is considered as an essential part of a natural lifestyle, and it is assumed that all diseases may be cured by *Urkost* diet and by a certain type of physical activity (*Ursport*). *Urkost*-style nutrition consists of raw vegetables, fruit, and wild herbs (such as dandelion). This raw food is neither washed nor cooked. Real followers of *Urkost* will not eat meat or any other animal products, but they are supposed to eat the insects that reside on plants which they eat (Konz, 1999). *Urkost* followers are also advised to eat humus from time to time.

The Internet is a quick and comfortable way of communicating between members of alternative medical movements, and this is also true of the *Urkost* movement. Its members use a web forum (http://www.urkostforum.de), in which they may search for and exchange information. This forum is targeted at people who practice *Urkost* or are interested in it. It is defined as a stage where *Urkost* is presented as a way of achieving a healthy lifestyle. The introduction page of the web forum explains that insults will be deleted and that a "loving and respectful" atmosphere of communication is desired. Brigitte Rondholz ("BR" in the following) – who is presented as the deputy of Franz Konz – is introduced as the regular moderator of the forum. She also appears to be the only administrator. This web forum provided the basic data for our hermeneutic analysis, which was conducted between April and September 2010. The users that come together on this web forum form a group that represents what Lave and Wenger (1991; Wenger, 1998) have described as a "community of practice". The users of the *Urkost* forum share an identity that is defined by a common interest in a certain subject. Even though most people would probably not consider the *Urkost* idea as a topic worth knowing anything about, *Urkost* followers obviously learn from each other and appreciate their shared competence. The members of the forum are strongly committed to their domain and engage in discussions and collective activities. They support each other and exchange information. The members of the *Urkost* community are literally practitioners and, as will be described below, they share and develop a collection of resources, such as stories, experiences, or routines for dealing with recurring problems.

Hermeneutic Sociology of Knowledge

The analysis presented here examined how knowledge is developed in the *Urkost* community, why its followers believe in this health knowledge – despite all logical inconsistencies and missing evidence –, and how this faith is accomplished collectively. What occurs here is a social construction of reality (cf. Berger & Luckmann, 1966), as people interacting with each other will experience that they may impute their own construction of reality and their corresponding emotions to others. Individuals will assume that what they think and feel would be correct because others seem to think similarly and report constructions that may be connected to one's own opinions. Such a collective affect-logic (cf. Ciompi, 1991) is not necessarily connected to individual processes of reflection. The experience that another person to whom an individual feels some emotional commitment takes a similar stand towards a certain topic appears to be sufficient for conveying to that individual the impression that a statement or assumption is true.

In order to conduct a qualitative analysis of the social construction of reality and the development of knowledge in the *Urkost* forum, we followed the hermeneutic sociology of knowledge. This approach is a complex theoretical and methodological concept that aims at reconstructing any type of interaction (Reichertz, 2004). An interpretation of data in the sense of this approach does not only consist of describing observations or reconstructing subjective meaning, but rather aims at detecting the inter-subjective meaning of actions. According to hermeneutic sociology of knowledge, the epistemologically appropriate attitude is one of general skepticism concerning social matters-of-course, and skepticism concerning prejudice in their interpretation. For our present analysis of knowledge communication and knowledge creation in the *Urkost* forum, we adopted such a hermeneutical attitude. The data that were used for our analysis are non-standardized data that occurred in a natural setting. They were not created for research purposes, so the collection of these data was not biased by any subjective observation schema. Moreover, our analysis was not restricted to previously determined research topics, but explored – in a first step that took several weeks – all contributions in this web forum, with respect to content, topics addressed, language, etc. This openness gave us various options of access and a bandwidth of topics from which we could select, as a second step, the most interesting and relevant modalities of knowledge construction in this community of practice. In this process of selection we made sure that the quantity of data obtained was sufficient for re-checking the explored aspects. In a third step, an analysis of the data was performed in such a way that contributions could be examined in more detail in their context. The insights that resulted from this procedure were challenged once again by amendatory and corrective interpretation by several researchers. Finally, the findings were 'compacted', transferred into written form, and illustrated with pertinent quotations from the forum (that represent only few examples from an abundance of data).

Analysis of Knowledge Construction

In the following four sub-sections, an analysis of the modalities of knowledge construction in the *Urkost* community will be presented. This analysis elicited four relevant aspects: impartment of a certain world view; role structure within the community; persuasion and reinforcement of *Urkost* followers; and their separation from standard medicine and people of a "different faith".

"*Let it be your green bible*" – View of the World

In the following, we will explore how *Urkost* influences its followers' way of life, at least according to discussions in this web forum. Our exploration provides an analysis that will be more concrete in the subsequent sub-sections on *role structure*, *persuasion*, and *separation*. When reading through the topics of the forum, one can easily see that nutrition and health do indeed play a central role, but they are by no means the only content. The 'right' attitude toward life and nature, to animal protection, religion, society, politics, sexuality, and many other topics are discussed as well. User "Waldelfe" wrote, for example (translated from German by the authors):

"As a conversion to Urkost changes nearly everything (for the better), sexuality will change as well. Nowadays, everything is about sex […] Especially young people […] attach great value to having a lot of sex, and finish a relationship when they do not turn on each other any more after some time. Franz Konz writes in his book that it is largely meat that leads to this excessive (?) craving and Schlechtkost[1] in general".[2]

Urkost appears to be a general attitude with a strong impact on all spheres of life, and carries a certain view of the world. This attitude appears to come along with criticism of society and culture. The reference to Konz' book as a guideline is another characteristic which can be found in many other contributions as well. The supposedly deficient lifestyle of the general population is attributed to eating *Schlechtkost*, i.e. "bad food". The discussion that followed this contribution centered on what characterizes good sex in terms of *Urkost*. There were many judgments concerning good or bad, right or wrong. Many norms and attitudes of society were questioned fundamentally and rejected, while *Urkost*-style attitudes were praised. In addition, the *Urkost* supporters propagated in their postings a life in primordial conditions in paradise. *Urkost* is supposed to lead to living longer, becoming or staying healthy, being sporty, strong and attractive, being happy etc. The users do not tire of listing all the advantages of *Urkost*. Those who live on an *Urkost* diet are supposed to reach an old age happily and healthily, as may be seen from a contribution by the user "Taraxacum":

"[…] here are my 20 reasons. 1 To become and to stay healthy. […] 19. Urkost is the ideal way. 20. To die of old age with at least 120 years." (Contribution date – format: dd.mm.yyyy: 07.02.2010, 03:32 pm).

People who live on *Urkost* commit themselves to a universal ideology that provides a guidance system and seems to bear some parareligious markings. The web forum was analyzed for religious aspects both linguistically and with regard to content:

1. The book by Franz Konz that was mentioned before is entitled *Der Große Gesundheits-Konz* (GGK) ("Konz's Great Health Book"), but some *Urkost* followers also call it the *Green Bible*. This definitive book is cited repeatedly and serves as an orientation and source of arguments. For example, user "Karlina" wrote:

   "You may find that all in the GKK, let it be your green bible, there you can find all the advice you need." (11.03.2010, 10:27 am).

   The content of this "bible" seems to be taken for granted, and citing from it is considered good form.

2. Recurring to biblical statements and Christian traditions may be found quite frequently (despite an ostensible denial of religiousness – as religion is related to society and culture and their negative connotations; instead, *Urkost* followers believe in "nature"). The following quotations by BR (a & b) and "Robert" (c), may illustrate this observation:
   a. *"May this forum […] be blessed."* (02.07.2008, 09:06 pm).
   b. *"I mean no harm to anybody, I even carry astonishingly much of this 'love thy enemy' with me that was demanded by Jesus."* (16.06.2010, 09:07 am).
   c. *"Paradise as described by the bible is not a utopian phantasy [...]"* (03.08.2010, 12:55 am).

   In one posting BR presented nearly messianic promises in biblical terminology:

   "Once we are a majority with our lifestyle […], then there will really be something for everybody […] and we can enjoy the abundance. […] Just like paradise." (20.05.2010, 10:23 am).

3. User "Urtica" wrote:

   "In our allegedly well-informed world, two superstitions appear to be ineradicable, because people want to believe in them: that they were capable of living on after death, and that there was some miraculous cure to take away their sins committed against life and nature. […] Believing in medical miracles is a global disease, one of mankind's many superstitions […]" (04.06.2010, 08:08 pm).

   While placing medicine at the same level as civilization and faith, this user employs Christian terms such as *sins* or *life after death.* Everything that is not in line with *Urkost* is attributed to superstition.

In addition to such religious aspects, it may be noted that the *Urkost* movement fulfils some criteria of a sect. The community is clearly focused on a leading figure (see the following sub-section on 'role structure'): BR is a leading figure of the *Urkost* movement, as may be seen from her great activity in the forum. She is successful in committing followers to herself and her doctrine of salvation (see sub-section on 'persuasion and reinforcement'). Salvation is equalized with healing from some illness, something reported quite frequently in the forum (e.g. *"Due to Urkost, we became healthy again"*). The community members perceive themselves as surrounded by enemies. The *Urkost* movement turns against *Schlechtköstler* ("bad food eaters") and considers them enemies. Criticism is not possible, neither from inside nor from outside; those who question certain positions, express criticism, or even support methods used by standard medicine will be outcast or vilified (see sub-section on 'separation from dissidents'). The *Urkost* movement is not a sect in the narrow sense: it does not persecute renegades (at least, there is no evidence in the forum) or expose dissidents, but it features some criteria of a sect. What we may conclude from this analysis is that knowledge creation in the *Urkost* community goes beyond questions of nutrition and health; it rather aims at providing an ideological guidance system.

"You are under special surveillance now" – Role Structure

How does the role structure in the *Urkost* community influence knowledge exchange between its members? Participants of the web forum differ with respect to the number of their postings. Apart from BR as administrator and a number of rather active members, there are also many registered users who are less active or have only participated for a short time. In order to be allowed to make contributions to the forum, one has to send an application to BR and needs to be approved by her as a contributor. At the formal level, we can only distinguish between the administrator and ordinary members. BR stimulates debates by introducing new topics into the forum. Even though BR is not always the first who raises a point for discussion, she gives her opinion on virtually every topic, so, for example, in a discussion on sexuality:

"You simply have to know that the mass media present sexuality in an absolutely incorrect way." (13.06.2010, 08:17 pm).

Very often it seems as if the members wait until BR expresses her opinion, in order to simply express confirming views. One gets the impression that BR usually has the final say and that she does not only play the role of an expert (due to her experience and knowledge), but rather the role of a guru. She also rebukes other users for stating "wrong" opinions. For example, in a discussion on whether it is 'correct' to eat maggots in fruits, user "Dawn" (who was a member for only two weeks at this time) had written:

"[…] I support natural and organic as well, but I'd prefer chemistry to eating maggots." (21.07.2010, 08:03 am).

BR replied:

"Prefer chemistry to maggots? Sorry, but you don't understand much of it yet, do you? […] Have a nice time in the Urkost forum, but you are under special surveillance now, honey. But just reading what the others write is nice, too, and instructional above all, right? " (21.07.2010, 11:59 am).

This reprehension seemed to result from a position of strength and showed that BR clearly takes on a role in which she decides what is right or wrong. And the other users accept BR's superiority: for example, user "orchidee" had written that in her opinion *Urkost* should not be practiced too strictly and that eating grain should be okay. BR answered by commenting every single sentence. Referring to the grain issue she wrote:

"[…] Humans are biologically not designed for grain. For every class of creatures on this planet nature has provided appropriate food. Each deviation leads to all sorts of problems: tumors, psychoses, depressions, allergies, etc." (26.03.2010, 11:49 am).

Subsequently, user "orchidee" confirmed:

"You were right, once again […] so far you have been right in the end in all discussions." (26.03.2010, 01:41 pm).

Finally, BR answered:

"I am happy and I enjoy watching you learning" (27.03.2010, 12:41 pm).

BR plays the part of a guru, and it becomes apparent that two goals are associated with this role: followers and potential new members are to be persuaded and kept (this will be addressed in the following sub-section on 'persuasion') and separation from dissidents (cf. sub-section on 'separation').

"You are on the right path" – Persuasion and Reinforcement

In this sub-section, we analyze how followers are reinforced and persuaded to living on *Urkost* prescriptions. On the one hand, we describe how the community communicates with new members and how knowledge is passed on to them. On the other hand, we analyze how skepticism in their own ranks is dealt with (handling of criticism from outside will be analyzed in the following sub-section). An *Urkost* follower is expected to take a clear stand on all kinds of issues – breast feeding, vaccination, natural contraception, or on whether vegans are allowed to kill ticks. This may be a great challenge to new members, and they tend to seek support from established *Urkost* followers. This is why it is interesting to analyze how new members are attracted. At first sight, the *Urkost* forum appears to be an open community. New members are welcomed to get to know the *Urkost* lifestyle. In order to analyze contributions that refer to attracting members, we selectively searched for responses to contributions by new users (newcomers were defined as users who joined less than three months ago). It was found that newcomers are received very cordially. New user "januschka" got three welcome greetings within one day. The newcomers' questions are taken seriously and answered in a personal way. For example, a new user was unhappy about eupepsia problems since he switched to raw food. He was praised for switching to raw food and his problems were traced back to being not consequent enough, as user "apfel" noted:

"You are on the right path, keep going […], but there is a lack of wild plants […] you eat too many cultivated plants" (05.09.2010, 10:10 pm).

Urkost followers try to give others a feeling of security. In this pursuit, a tough-minded attitude towards *Schlechtkost* is given high priority. Usually, these users refer to some personal experience, but factual arguments are rare. Negative presentations of *Schlechtkost* are contrasted with salutary *Urkost*. This contrast is the basis for reinforcing and motivating followers. When users describe some conflict with non-*Urkost* people, they experience strong support in the forum: they are incited not to listen to them, and supplied with encouraging stories and phrases that can be used as replies in debates. But as soon as a certain period of grace is over, it turns out that the users monitor each other's compliance with the *Urkost* regulations, and evaluate each other accordingly. Members who practice *Urkost* in a half-hearted way, will experience harsh criticism (see above). Sometimes this results in the withdrawal of a member. As an example, we analyzed a dispute with user "Petra". The original debate cannot be depicted completely, as this user withdrew from the forum and her contributions were deleted. But we may reconstruct some of her statements as they have been quoted by other users. "Petra" said goodbye with these words:

"[…] Even though […] I have met some lovely people, I found out very quickly, unfortunately, that one's own opinions are only accepted if they are consistent with those of the person who runs this forum."

She also complained about *"insults below the belt*" and concludes:

"I can only advise against fully believing in ideologies, health literature, or scientific reports."

BR reacted with disapproval. Other members joined in and attacked "Petra". User "amelia" wrote:

"All I can say about Petra: She really hasn't understood anything. Moreover, she lets herself be influenced by propaganda against the Urkost forum and against you [BR], instead of thinking for herself." (26.06.2010, 10:54 am).

User "cherrycurry" added laconically:

"Eating meat makes you daft, it seems" (26.06.2010, 11:52 am).

BR also likes to impute "*shabbiness of character*" to dissenters (20.06.2010, 11:52 am).

Reinforcement, motivation, and mutual support occur at an emotional level and are based on a separation from skeptics in the forum users' own ranks (such as "Petra") and dissidents (cf. next sub-section). Many discussions are rather personal, and this may have a positive effect on members by inducing a feeling of security. But this personal atmosphere of communication may also be provocative and offensive. At the surface the healing effects of *Urkost* are praised, but underneath many members seem to be rather vulnerable. The *Urkost* followers appear to be easily assailable and under constant pressure to justify. Critical dialogue is hardly imaginable from such a position and not really taking place. Discussions are dominated by references to personal experience. The goal is not to exchange valid arguments, but to be there emotionally for the like-minded and to signal that one is on the right track.

"Let them die stupidly from diseases" – Separation from Dissidents

The followers consider *Urkost* as the only right path. Accordingly, people of some "different faith" are despised, or pitied at best. When conflicts with *Schlechtkost* eaters are discussed, many statements express the superiority of *Urkost* followers in terms of a principle of selection. For example, user "Irmisato" wrote:

*"Don't listen to those people who want to put you off Urkost. Just say: 'We'll talk again in 50 years'. And think: '… in case **you** are still alive then'."* (10.06.2010, 01:35 pm).

This principle of selection (in terms of social-Darwinistic ideology) is even more apparent in a comment by user "Robert":

"Let them die stupidly from diseases, nature is tough sometimes, but always wise…" (10.06.2010, 05:11 pm).

What finds expression here is the notion, also known from other pre-modern views of the world, that one is struck by illness as a consequence of guilt or sin. This was a common attitude in the ancient world and the Middle Ages (Riese, 1953). The members of this forum appear to take for granted that the *Urkost* lifestyle is superior, and that this is expressed by their own health. They believe that this superiority will be evident in the long run and find expression in a longer life of *Urkost* followers. An explanation for the supposed misbelief of dissidents is provided be user "Nera":

"The majority still believes in the lies of science and does not make an effort to question them. Authorities mean more to them than critical thinking and insight." (13.06.2010, 08:13 am).

This statement not only contains an explanation for the behavior of dissidents, it also implies a superiority of *Urkost* followers in thinking and reasoning: *Urkost* followers consider to distinguish themselves from non-believers by greater intellectual ability as well. In order to be separated from dissidents more easily and protected from criticism, it is sometimes discussed to switch to a closed web forum. But this proposal was rejected by user "Ritchen":

"By no means should the entire Urkost forum become private. As you [BR] had said before, this is VERY IMPORTANT because of public relations in particular. We want to tell people that this road exists. And many have already found it – through Brigitte, her forum, her blog, her seminars." (20.06.2010, 12:17 pm).

But it appears that an additional closed forum is also in existence, in which even more radical ideas may be exchanged. There is a posting by BR in which she mentioned a "private *Urkost* forum" and explained:

"By all means, one can be more explicit there than here, and I will permit access only restrictively to hand-picked people." (01.07.2010, 04:58 pm).

As far as standard medicine is concerned, the attitude expressed is critical, to put it mildly. There is fundamental doubt that standard medicine is able to cure any disease at all. For example, user "Robert" wrote:

"Who cures is right! And somebody has to prove to me that physicians have ever cured anything with their standard medicine!" (10.06.2010, 05:11 pm).

Another example is a contribution by user "apfel":

"She [BR] is a fighter for the good on all levels (Not only when it is about the lies of standard medicine, woman-destroying contraceptive pills, vaccination, human suffering)." (31.07.2010, 05:17 pm).

This, generally, very critical attitude also applies to specific procedures of standard medicine, for example, vaccination, as can be seen from a contribution by user "Jana" in a discussion on the necessity of inoculation and the normal number of infectious diseases:

"We only have one infection per year. After two years, still breast fed and completely unvaccinated. My personal assumption is that the main cause for this alarming state of the children's health is vaccination." (31.07.2010, 03:49 pm).

Other users took the same line, for example "Nera" and "apfel":

*"Nice, if you can lay the blame for secondary complications of vaccination on some phantom germs. Saves jobs, doesn't it? *cough*"* (31.07.2010, 03:56 pm).

"Vaccinations are really very, necessary/important … for the moneybag of the greedy pharmaceutical industry. It is sad that innocent children are exposed to dangerous vaccinations." (31.07.2010, 01:41 pm).

The sarcastic tone in these comments is characteristic of contributions that deal with standard medicine. It seems as if this diction is supposed to express their authors' perceived own superiority, or, perhaps, resignation in the face of the incorrigibility of the rest of the world. Critical discussion on the necessity and the risks of vaccination are, of course, not restricted to groups like the *Urkost* community, even though the particular way in which this fraternity addresses the topic is rather exceptional. What is even more peculiar is their attitude to HIV/AIDS, where the discrepancy between standard medicine and *Urkost* views is particularly apparent. The *Urkost* followers not only criticize the methods of HIV/AIDS treatment, but their criticism goes further: they doubt the existence of AIDS, as a viral disease, in general. User "Nera", for example, claimed:

"[…] Aids is not a disease caused by a "virus", and it is curable at any time." (10.07.2010, 12:31 pm).

This statement was broadly supported by other users. User "Robert" wrote on this topic:

*"Correct. Aids is the outbreak of various diseases due to a f***ed up immunity system. Earth fast*[3]*, Urkost, sports and it can be cured."* (20.07.2010, 11:29 am).

AIDS is considered as a lie that is maintained due to financial interests. At someone's suggestion to move this topic of AIDS healing to a private forum, BR answered:

"The Aids critics are reputable and there is no reason to hide them. Rather those people should hide who keep the money-printing machine running although they know better." (21.07.2010, 07:04 pm).

So far, we have analyzed a rather theoretical attitude to standard medicine and its representatives. But what will *Urkost* followers do when they meet physicians, who represent standard medicine, and how will they react to diagnoses by conventional medicine in their everyday life? Some forum contributions give the impression that some *Urkost* followers indeed turn to doctors of medicine – especially with severe diseases, such as tumors, or when children are concerned. The recommended treatments, however, are often rejected. Instead, they try to cure themselves the *Urkost* way, as can be seen from the example of user "Jana" who reports that she suffered from a tumor in the parotid gland:

"[…] Surgery denied. Switched to raw food the next day, after 6 weeks tumor not palpable anymore. In April 2010 finally another MRI, and last week finally its results. Tumor completely disappeared." (01.06.2010, 09:07 pm).

What is interesting here is that methods of standard medicine such as an MRI are accepted and taken for granted, and that the results of this procedure are awaited eagerly, as can be concluded from the double use of the word "finally". Further contributions in this thread show that the other users do not only regard this as a fight against the tumor, but also praise it as a victory against the physicians. User "Bartok" wrote in this context:

"Hello Jana, that is great You can be veeeery proud of you that you showed the ropes to the tumor and to the doctors. What did the doctors say about it? Would be nice if you told us how they responded to the refusal of surgery and the disappearence of the tumor." (02.06.2010, 07:57 am).

"Jana" answered that the doctors would deny any relation between *Urkost* and the healing, and expressed her disappointment:

"I would have been glad if I could have made the doctor think. But this was not the case." (02.06.2010, 10:28 am).

This disappointment reflects the ambivalent relationship to physicians, which was mentioned above. On the one hand, as representatives of standard medicine, they are considered as enemies, in theory. But, on the other hand, they are consulted, their diagnoses are regarded as real, and their evidence for diseases and for recovery are completely accepted. In the stage of treatment, however, what they say is irrelevant. It seems as if doctors were important authorities, whose persuasion would be a great success. Physicians and their diagnoses are by no means ignored, they rather represent a point of friction that is heavily discussed.

An aspect on which the forum is not very explicit is the attitude of *Urkost* to other alternative medical approaches and procedures. One example in which another alternative medical procedure was advised was the case of a mother (user "Kleeblatt2007") who wrote that her little son had started to be cross-eyed. On the one hand, she was surprised:

"We are sad, because we had thought that this would not occur with Urkost" (13.07.2009, 08:31 pm).

But she admitted that she had eaten meat from time to time during pregnancy, and assumed that this might be a potential reason for her son's strabismus. She asked for suggestions, and various alternative medical procedures were recommended to her which are not based on Urkost. But this seemed to be no problem here. Users "alice", "Mangofan", and "uma" recommended various methods:

"I finally came across Glenn Doman […] I would never recommend glasses. Because these do not stimulate the brain, quite the opposite." (14.07.2009, 12:15 am).

"Strabismus is not a disease but a dysfunction that can be resolved in many cases, following Dr. Bates in his book […]" (17.07.2009, 10:35 am).

"I started a cranio-sacral-therapy with her [her cross-eyed daughter] […]". (30.09.2009, 11.08 am).

A separation can be found in this discussion as well, but, again, mainly from standard medicine, which can be seen in some side blows, for example in a contribution by user "Mangofan":

"Ametropia is always caused by stress, because the external muscles of the eye disarrange the eyeball, especially when a toddler is examined by an eye doctor." (17.07.2009, 10:35 am).

Surprisingly, the professional medical background of authorities is considered as positive if the approaches of these medical authorities are like-minded with the respective user's view of the world. Two examples were provided by the users "alice" and "Mangofan":

"Glenn Doman […] is a physiotherapist and has worked for more than 60 years with brain-injured children." (14.07.2009, 12:15 am).

"Fortunately, Dr. W. H. Bates has been doing research for nearly 40 years and found that eyesight improved with anatomically correct use of the eyes with the brain. He was a physician, ophthalmologist, and eye surgeon." (21.07.2009, 06:28 am).

And even BR applies this strategy, for example in the following posting:

"'Humans are not really granivores', says Karl Pirlet, Professor of Internal Medicine. only for the last 15,000 years have humans eaten grain, too short for a real genetic adaptation." (26.03.2010, 11:49, am).

This is an indication of what might be called the *fundamentalist eclecticism* in the *Urkost* community: anecdotic knowledge is indiscriminately mingled with medical and scientific findings. Often, scientific medicine is fundamentally criticized in the same breath as a reference to some scientific study is cited as evidence for the trueness of one's own world-view.

Conclusion

In this paper, we provided a qualitative hermeneutic analysis of a web forum, in order to examine how a (rather radical) alternative medical community of practice, the so-called *Urkost* movement, applies the Internet as a platform for knowledge exchange and knowledge construction. The *Urkost* community is concerned, primarily and at first sight, with aspects of nutrition and health, but what happened in this forum exceeded the construction of health knowledge and addressed an ideological view of the world in general. Even though a holistic approach is typical of alternative medical communities, the way in which a construction of reality took place in the *Urkost* movement was definitely exceptional. So our findings cannot be generalized to knowledge construction in other alternative medical communities, but it was still revealing to elaborate on the structures of knowledge exchange that emerge from our hermeneutic analysis of this particular alternative medical fraternity. Even though the exchange processes that we analyzed did not really represent knowledge development in scientific terms, what occurred from the *Urkost* point of view was indeed a form of knowledge advancement, users learning from each other, forming a collective identity, and developing a common repertoire of resources. Our hermeneutic procedure identified four aspects of knowledge creation within this community of practice.

Firstly, we examined to what extent *Urkost* is more than just alternative medicine to its followers. It appears that *Urkost* is rather a holistic ideology and attitude towards life adopted by members of the *Urkost* forum. Their contributions showed that they do not only refer to curing diseases, but address a great variety of topics of everyday life. The influence of *Urkost* on these people's lives bears some parareligious markings. Concerning the second aspect of knowledge creation, the role structure in the forum, it was found that there is one administrator who clearly took the position of a leader or guru. Such an outstanding position is actually untypical of a community of practice, but its existence is comprehensible – and this is in line with theory – because this person plays a significant role when it comes to ensuring compliance with group norms and the formation of identity: she initiated new discussion threads, commented other contributions very frequently, and was always on the spot when it came to defending the *Urkost* lifestyle in a quite dominant way. The third part dealt with the question how communication was used to put newcomers on the right track and rebuke members who expressed deviant opinions. It was found that the *Urkost* forum welcomed new members in a quite cordial and open way. Critical questions or comments, however, were not accepted, and answered brusquely and reprehensively. Both the motivation of the forum members and responses to critics were communicated in an emotional and personal manner. The final part of the analysis of knowledge-creation structures examined the attitude of the *Urkost* community to standard medicine. We found a critical, even hostile communication style when the topic of doctors, their treatments, and standard medicine in general came up. The behavior of standard medicine representatives was criticized and it was even insinuated that conventional medicine would do harm to patients intentionally and out of greed. By clear-cut communicative separation from standard medicine, a feeling of supremacy and, perhaps, also a sense of uncertainty of the *Urkost* followers were expressed.

It may be concluded that the community that was examined here is, in large parts, a prime example of a community of practice: the identity of this group is defined by a shared interest, its members are engaged in joint discussions, and they have developed a shared repertoire of stories and cases for their practice. Many things that were going on in this community may appear absurd or insane to outsiders, but it makes perfect sense – from a systems theoretical, constructivist point of view: the social system (in terms of Luhmann, 1995) described here uses its own operations in order to distinguish itself from its environment (e.g. from standard medicine) and, in doing so, determines what is accepted as legitimate content of its own communication. As a consequence, communication takes place in a way that allows connectivity for further pertinent communication. Within the scope of this self-organization (in terms of autopoiesis) the system maintains itself, reduces complexity from the environment and constitutes meaning (cf. Kimmerle, Moskaliuk, Cress, & Thiel, 2011).

Endnotes

(1) The term *"Schlechtkost"* – like *"Urkost"*, not really an existing word in German – is often used by the *Urkost* followers and may be translated as "bad food". This neologism is reminiscent of the terminology used in G. Orwell's "1984".

(2) No contribution date is provided for this quotation, because this user left the web forum early in September 2010 and all her contributions were deleted. We had taken down this quotation at an earlier session, but could not find it any more in the web forum when we wanted to note the contribution date.

(3) *"Erdfasten"* – literally "earth fast" – is the *Urkost* "starting diet" that people are supposed to go through before they start living on *Urkost*.

References

Berger, P. L., & Luckmann, T. (1966). *The social construction of reality: A treatise in the sociology of knowledge*. Garden City, NY: Anchor Books.

Bratman, S. (1997). *The alternative medicine sourcebook: A realistic evaluation of alternative healing methods.* Los Angeles: Lowell House.

Campbell, C. (1972). The cult, the cultic milieu and secularization. In M. Hill (Ed.), *A sociological yearbook of religion in Britain* (Vol. 5, pp. 119-136). London: SCM Press.

Ciompi, L. (1991). Affects as central organising and integrating factors. A new psychosocial/biological model of the psyche. *The British Journal of Psychiatry, 159*, 97-105.

Kimmerle, J., Cress, U., & Held, C. (2010). The interplay between individual and collective knowledge: Technologies for organisational learning and knowledge building. *Knowledge Management Research and Practice, 8,* 33-44.

Kimmerle, J., Moskaliuk, J., Cress, U., & Thiel, A. (2011). A systems theoretical approach to online knowledge building. *AI & Society, 26*, 49-60.

Konz, F. (1999). *Der große Gesundheits-Konz*. München: Universitas.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.

Luhmann, N. (1995). *Social systems*. Stanford: Stanford University Press.

Reichertz, J (2004). Objective hermeneutics and hermeneutic sociology of knowledge. In U. Flick, E. v. Kardorff, & I. Steinke. (Eds.), *A companion to qualitative research* (pp. 290-295). London: Sage.

Riese, W. (1953). *The conception of disease: Its history, its versions and its nature.* New York: Philosophical Library.

Salmon, J. W. (1984). *Alternative medicine: Popular and policy perspectives*. New York: Tavistock.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge: Cambridge University Press.

Learning Science through Knowledge-Building and Augmented Reality in Museums

Susan A. Yoon, University of Pennsylvania, Philadelphia, PA, yoonsa@gse.upenn.edu
Karen J. Elinich, The Franklin Institute, Philadelphia, PA, kelinich@fi.edu
Christopher Steinmeier, Joyce Wang, Sean Tucker, University of Pennsylvania, Philadelphia, PA
Email: cst@dolphin.upenn.edu, joyce.s.wang@gmail.com, tuckers@med.upenn.edu

Abstract: Although learning science in non-school settings has shown great promise in terms of increasing interest and engagement, few studies have systematically investigated and produced evidence of improved cognitive gains and higher order inquiry skills. Furthermore, little is known about what and how students learn through the digital technologies that are increasingly being used in these non-school settings. Through an experimental design, this study compared four conditions for learning science concepts with augmented reality technologies in a science museum. Using knowledge-building scaffolds known to be successful in formal classrooms, we hypothesized that students would show greater cognitive gains and higher order inquiry skills than when scaffolds were not used. Through the use of digital augmentations, the study also provided information about how such technologies impact learning in informal environments.

Introduction

The issue of declining participation by America's youth in science, technology, engineering, and math (STEM) education and careers has received a great deal of attention over the last few years from industry, government, and education sectors (Business Roundtable, 2005; Domestic Policy Council, 2006; U.S. Department of Education, 2007). Experts predict that the availability of STEM jobs will continue to increase such that, by 2012, the number of positions in science and engineering will have outpaced the number of qualified people to fill them by 26% and 15% respectively (NSF, 2006). Additional pressures on workforce development exist with the emphasis on 21st century skills (Partnership for 21st Century Skills, 2007) by which learning from and using digital technologies both for developing conceptual knowledge and process skills must figure prominently in education. As new scientific and technological developments continue to impact people's lives, there is further interest in providing educational opportunities that will increase general levels of scientific and technological literacy. Given this enormous need, there is increasing recognition that STEM education can no longer be the sole responsibility of formal schooling. The recent National Research Council (NRC) report on learning science in informal environments (Bell et al., 2009) examines the potential that non-school settings such as zoos, aquaria, and museums have for engaging large portions of the population in real-world scientific investigation. Furthermore, the report notes, where NCLB requirements have limited the amount of in-school science instructional time, informal programs actually serve as essential venues for learning. However, the report and others (Rennie et al., 2003) also highlight the need for systematic studies of learning designs in order to realize the potential for the field of informal science to contribute to STEM education and career development.

This study investigated three related critical gaps in understanding of informal learning, as outlined in the NRC report. First, while there is ample evidence of increased levels of interest and engagement, evidence for improved cognitive gains is less convincing. This is partially due to the free choice, episodic structure of activities characteristic of informal environments. Second, as more educational technologies are being used to assist in the development of conceptual knowledge, little is known about how digital platforms improve the learning experience in these settings. Finally, while designed interactive activities have been shown to increase important scientific skills—such as manipulating and observing—higher order inquiry skills—such as reflection, making predictions, drawing conclusions, constructing generalizations, and argumentation—are less frequently demonstrated. This study considered these critical gaps together by investigating the effects of digital platforms in an informal environment on cognitive gains, with particular emphasis on higher order inquiry skills. Using knowledge-building scaffolds for peer-to-peer discursive interaction and collective knowledge construction (Scardamalia, 2002), the study aimed to determine whether and how learning through digital platforms might be enhanced through a knowledge-building design, which had not previously been applied in informal environments. The research was conducted at a premiere science museum in a large urban city in northeast USA using augmented reality visualization technologies. The specific questions investigated were: 1) In what ways do digitally augmented visualizations of scientific phenomena assist learners in developing conceptual understanding in a museum environment? 2) To what extent do knowledge-building scaffolds added to visualizations improve higher order inquiry skills? and 3) How can knowledge-building scaffolds, that are highly successful in sustained and continuous classroom activities, be adapted to support single-visit episodic and free choice engagement in museums?

Theoretical Considerations

The following three areas of research in STEM education and the learning sciences informed the study's goals. *Digital Augmentations and Visualizations*: Augmented reality applications or digital augmentations of real-world phenomena have increased in formal education and other knowledge domains such as medicine over the last few years (John & Lim, 2007; Klopfer & Squire, 2008). Digital projections are programmed to enhance an object or environment in order to simulate or visualize information that is not normally observed. For example, in one game called *Environmental Detectives*, students can access virtual underground water sources to determine impacts of a toxic spill (Klopfer & Squire, 2008). Visualization tools in science education research have been found to produce positive learning gains such as challenging and correcting misconceptions (Tasker & Dalton, 2008), concretizing abstract theories (Dori & Belcher, 2005), and understanding scale of agents within systems (Sengupta & Wilensky, 2009). Although one group of researchers found positive outcomes in terms of interest and engagement when using augmented reality devices in a museum setting (Waite et al., 2004), little is known about cognitive impacts. *Knowledge Building and Scaffolds*: With respect to improving conceptual understanding and promoting higher order inquiry skills, one pioneering program of research in the learning sciences involves computer-supported intentional learning environments (Hewitt & Scardamalia, 1998; Scardamalia & Bereiter, 1994). Both the technological application and associated pedagogy use educational scaffolds to enable public, collective contributions that shape the knowledge constructed in the learning community. Such scaffolds include prompts for consensus building, generalizations, differentiation between evidence and theories, peer evaluation, and argumentation. For example, a prompt such as "*My theory is...*" encourages students to use evidence to construct a more general understanding of a class of scientific phenomena. Similarly, the act of creating a "rise above" note provides students with opportunities to think across diverse ideas and to offer conclusions about how the collective community views a scientific issue such as the benefits of genetic engineering (Yoon, 2008). *Learning in Informal Environments*: As noted in the NRC report, learning in informal spaces is fluid, sporadic, and participant driven—characteristics that are thought to increase engagement over highly structured formal classroom experiences. Often learning activities occur in single-visit episodes (Falk & Dierking, 1992) where individuals learn on their own with little follow-up or reflection. To date, knowledge-building programs have been used in formal classroom environments where there are opportunities for extended, recurring investigations—opportunities that informal environments generally lack. However, there is some evidence from museum visitor studies, that significantly greater cognitive gains can be achieved when objects are accompanied by interpretive labels (a simple form of scaffolding) than when there is no structure provided (Allen, 1997; Borun & Miller, 1980), suggesting that common ground between the two environments may exist. These results, and the positive impact of knowledge-building activities when coupled with computational tools, provide the rationale for undertaking this study.

Methodology

Participants and Context

Study participants were recruited from 6th and 7th grade classes who attended the museum on a field trip. This grade band was chosen deliberately. By middle school, most children are developmentally able to theorize about scientific phenomena. Furthermore, the study's scientific content was electricity, which local standardized curriculum covers in grade 4. Thus, all students would have some prior knowledge to access. The classes came from three schools within a high needs urban school district with an average of 92% of students qualifying for free or reduced price lunch. 119 students participated, with a gender ratio of 55% female to 45% male. While this was not the first time these students had visited the museum, it was the first time they had been invited to participate in learning research with the museum's educators.

The study leveraged a currently funded large scale National Science Foundation informal science education project in which the central goal is to design, integrate, and increase the use of educational technologies within the museum experience. Several tools are under construction including digitally augmented exhibit devices. While the larger project focuses on the viability of using digital technologies in informal settings, the present study extended this focus by investigating what and how scientific concepts were learned through such applications. The investigation used a prototype exhibit device called "Be the Path" that illustrated electrical conductivity and circuits (see Figure 1). The students had no previous access to this device which consisted of two metal spheres on a table, approximately one foot apart, with one connected by a wire to a battery and the other connected to a light bulb. The students attempted different configurations to complete the circuit and light the bulb by touching the balls. Once the circuit was completed, a projected visualization of electrons flowing around the complete loop appeared. The instructions on the device provided little direction, simply suggesting "try to complete the circuit."

Students were grouped into four conditions such that each of the three participating schools had some students encountering each of the four conditions so that each condition analyzed below includes students from three schools. The conditions were constructed to represent increasing use of scaffolds for learning through digital augmentations and knowledge-building structures. Condition 1 (C1) served as the control group with no digital augmentations or knowledge-building structures. Condition 2 (C2) represented the device with the digital augmentation but no other scaffolding. This condition was designed to represent the average museum visitor who participates mainly through hands-on sensory experiences or trial and error (Bell et al., 2009). Condition 1 and 2 also directly address the first research question, which is to understand the impact of digital augmentation on conceptual understanding. Condition 3 (C3) featured some scaffolding in which the equivalent of labels in the form of directed questions were posted in the room for participants to reference. This condition was designed to represent the ideal scenario for exhibit designers as they provide labels, signs, and explanations with the expectation that visitors will read them. The posted questions were: What happens when you touched both metal spheres? When you touched only one? What happened to make the light bulb light up? What does the projection show? What are you supposed to learn by using this device? Condition 4 (C4) represented the condition in which both digital augmentation and knowledge-building applications were used. Students were instructed to work as a group of three. Each group was given the same questions as students in condition 3 and additionally asked to brainstorm possible answers, give reasons for each, decide as a group, and write their collective response on a worksheet. These directions were listed in a small box at the top of the worksheet. Other knowledge-building scaffolds were added to the worksheet questions in the form of directed prompts such as "*Our hypothesis is...*" and "*Our theory is...*" Students also had access to a bank of completed worksheets which were posted on the wall in order to evaluate their own ideas against others through two more knowledge-building scaffolds: "*Others have said...*" and "*We agree/disagree with them because...*" Conditions 3 and 4 were constructed to respond directly to the second two research questions, which investigated whether knowledge-building scaffolds increased higher order thinking and how a knowledge-building process might be adapted to informal environments. Students in conditions 1, 2, and 3 completed the worksheet individually after they finished playing with the device while students in condition 4 actually completed the worksheet during the experience.

Figure 1. Be the Path Device with Digital Augmentation.

Students were randomly assigned to each condition. As the study was embedded in a whole class trip to the museum, we wanted all students to participate in the activity. However, logistical factors subsequently caused variation in the number of participants analyzed in each condition. Some students lacked authorized consent to participate forms. Other students were absent for the pre-test data collection. See Table 1 for the number of participants in each condition.

Table 1: Number of participants in each condition.

Condition	Number of Participants
1	18
2	22
3	31
4	16 groups of 3 (48)

Data Sources and Analyses

Four data sets were collected and analyzed through a mixed-methods approach.

1. A conceptual knowledge survey was administered to students in each group before and after the intervention. The survey posed five general multiple choice questions related to the scientific topic of electrical conductivity and circuits such as, "Which class of elements best conducts electricity?" An open-ended question on the survey also solicited responses that demonstrated knowledge directly related to the device experience, i.e., "Think about an electric circuit that supplies electricity to a light bulb. What parts

make it work so that the bulb lights up?" These responses were coded on a five point Likert-scale from no understanding (0) to complete understanding (4). Collectively, the highest possible score on the conceptual knowledge survey was nine. A paired-samples t-test was conducted to determine whether there was a statistically significant gain in conceptual knowledge within each condition. A repeated measures ANOVA was conducted on the data set to determine whether there was a statistically significant difference between the mean gains.

2. With respect to worksheet responses, the last question, "What are you supposed to learn by using this device?" was intended to elicit responses that demonstrated generalized higher-order thinking, i.e., understanding of electrical circuits and how the human body functioned as a conductor in the device experience. Responses were scored on a four point Likert-scale from no understanding (0) to complete understanding (3). A response that demonstrated complete understanding satisfied the following criteria:

   Level 3 Description:
   - Student identified that the human body can complete a circuit to conduct electricity.
   - Student identified both concepts (completed circuit and conductivity).
   - Student demonstrated accurate understanding of both concepts, even if the terms "circuit" or "conduct" are not explicitly used.
   - Student correctly used simile or metaphor to explain concepts.
   - Even if student identified all components, if a part of the answer is wrong or contains a misconception, it cannot count as complete understanding.

   An ANOVA was conducted on the data set to determine whether there was a statistically significant difference in responses between the conditions and a post-hoc Tukey analysis was conducted to determine the source of the difference.

3. In order to investigate how knowledge-building scaffolds impacted the nature of the museum visit, 10 groups (30 students) in condition 4 were randomly selected for short interviews. Questions solicited responses from students to determine the utility of worksheet questions and the knowledge-building scaffolds in helping them learn about the scientific topic. Responses were summarized and used to identify which scaffolds were more or less effective. Responses of scaffold helpfulness were tallied and a chi-square test was conducted on the frequencies in each scaffold category.
4. The last data source integrated observation field notes and anecdotal reflections from the five researchers. These notes and reflections were used to triangulate findings from the other data sources and also to provide plausible rationales and further hypotheses that might explain the results we obtained.

For the open-ended questions on the conceptual knowledge survey and the worksheet, the data were qualitatively and systematically mined by the researchers (Strauss & Corbin, 1998). Codes were established, a categorization manual was constructed, two external graduate students were trained on the coding scheme, and inter-rater reliability was obtained on 20% of the data (35 mixed pre- and post-intervention open-ended responses and 18 worksheet responses). In both cases, greater than 90 percent agreement was obtained.

Analysis

Conceptual Knowledge Survey

Figure 2 shows the mean raw scores obtained for pre- and post-intervention conceptual knowledge surveys in all four conditions. Over all, across pre- and post-intervention surveys, the means were low, ranging between 2.6250 to 3.6818. Students in C2 (the condition with just the digital augmentation) had the greatest gains and the highest raw score on the survey.

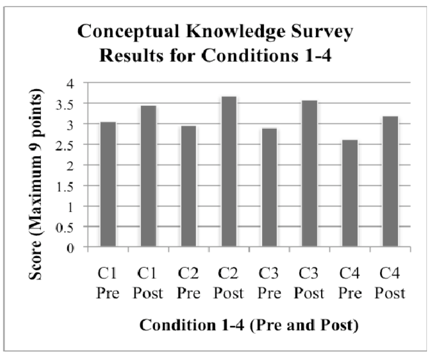


Figure 2. Means of Raw Scores for Pre- and Post-Intervention Conceptual Knowledge Surveys for Conditions 1-4.

Table 2 shows the results of paired-samples t-test conducted within each condition. The table shows that gains in all conditions except C1 were statistically significant. However, the repeated measures ANOVA showed no significant differences between the mean gains. From the observation field notes and anecdotal reflections, researchers identified some difficulties students had in referring to and completing the worksheet questions and the post-surveys in C3 and C4. These students wanted to rush through the activity in order to participate in the rest of the field trip and appeared to put less effort into the post-intervention surveys.

Table 2: Results of paired-samples T-Test comparing means within conditions.

Condition	Mean Difference	SD	t	df	Sig. (2-tailed)
1	.38889	1.81947	.907	17	.377
2	.72727	1.07711	3.167	21	.005*
3	.67742	1.83280	2.058	30	.048*
4	.56250	1.41280	2.758	47	.008*

* $p<0.05$

Worksheet Responses

The analysis of worksheet responses that evaluated students' abilities to generalize about electrical circuits and conductivity yielded some positive results toward understanding the impact of the different study conditions. Figure 3 shows an increasing trend from C1 to C4 in the means obtained of generalized higher-order thinking. Out of a possible score of 3—which indicated a complete understanding—students in C4 scored the highest with a mean score of 2 (representing the category of Partial Understanding) while students in the remaining conditions scored between 1 (Little Understanding) and 1.5.

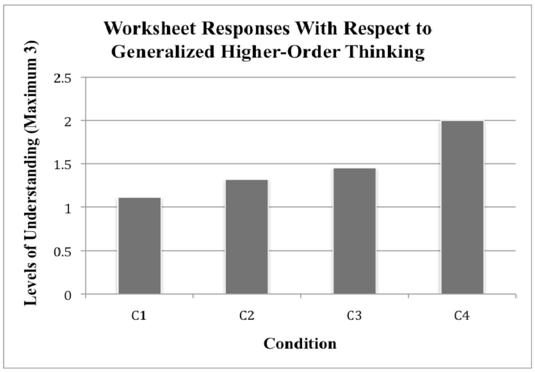


Figure 3. Mean Scores of the Generalized Higher-Order Thinking Analysis for Conditions 1-4.

An ANOVA showed a significant difference between the means of the four conditions, $F(3)=4.560$, $p=.005$. A post-hoc Tukey analysis showed that the difference was attributable to the higher mean of C4 which was significantly higher than C1 ($p=.004$) and C2 ($p=.028$), and marginally higher than C3 ($p=.077$). Sample responses from students in C4 along with their codes compared to responses from students in other conditions illustrated these results:

Student 1 (C4): I learn the body can act like wires to connect stuff and there are electric charges in the body. (score = 3)
Student 2 (C4): How our body transported electricity through the circuit into the metal spheres to make the bulb and images appear. (score = 3)
Student 3 (C2): You have electrical currents in your body. (score = 1)
Student 4 (C3): How energy flows through your body. (score = 2)

Interview Responses

In the small group interviews, condition 4 students were asked to provide responses about whether they believed the following five categories of worksheet and knowledge-building scaffolds were helpful: worksheet questions; directions on how to work in a collaborative group listed at the top of the worksheet; collaborating in a group; knowledge-building prompts; and a bank of answers from other groups. Figure 4 shows a graph of student responses in each of the categories. There was unanimous agreement that collaborating in a small group was helpful while the directions at the top of the worksheet were least often identified as helpful. Greater than 50 percent of the students said that the remaining three categories of scaffolds were helpful. When asked what they thought was the most and least helpful scaffold, 100 percent of the students identified collaborating in a group as most helpful. The least helpful scaffolds were identified as the knowledge-building prompts (57 percent) and the directions (37 percent). However, there is some evidence to suggest that students had difficulties understanding why the prompts were included. For example, one student said, "We didn't even know what they was about, like, but we just put an answer that others have said, like on the board." When students did understand the rationale behind including the prompts, their helpfulness increased. For example, one student answered, "Because we knew what we was supposed to give. Like the hypothesis, what we believe it is, and the theory is what we think it is." Other students said that they weren't accustomed to answering questions like that. Thus, responses to whether or not students felt the knowledge-building prompts were helpful appeared to be influenced by their familiarity or understanding of the prompt vocabulary such as "hypothesis" and "theory." Similarly, with the directions, some students stated that they didn't read them because they were just focused on answering the questions. The following interviewer and group exchange illustrated this point:

Interviewer: Alright, remember these directions in the box. OK? Did these directions help you figure out how to come up with the best answer?
Student #1: Didn't read them.
Interviewer: You didn't read them?
Student #1: Nope.
Student #2: I didn't see it.
Interviewer: What about you?
Student #3: I didn't read nothing.
Interviewer: You didn't read them. OK. So why did you guys not read this?
Student #1: I don't know, it's just like...
Student #2: 'Cause we was studying. We was focused on the questions.

Comments in the researcher field notes and anecdotal reflections support this finding. After a longer interaction with the device than the other three conditions, some students appeared to be anxious to complete the worksheets and had a "just get it done" disposition. Still, considering students in condition 4 showed greater ability to generalize from the experience (Figure 3), the worksheet and knowledge-building scaffolds seemed to provide some utility in terms of learning the content. A chi-squared analysis revealed a statistically significant difference between the frequencies found in the five categories, $x^2 = 9.96$, p=.0411.

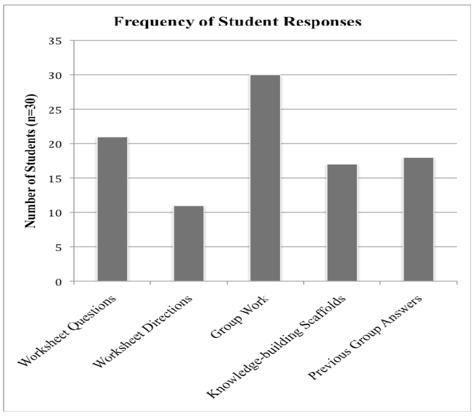


Figure 4. Frequency of Student Responses Indicating Helpfulness of Categories of Worksheet and Knowledge-Building Scaffolds.

Discussion

The explanation and implication of study findings are discussed in order of the three research questions.

1) In what ways do digitally augmented visualizations of scientific phenomena assist learners in developing conceptual understanding in a museum environment?

The results of the conceptual knowledge survey that tested for increases in general knowledge of electrical conductivity and circuits showed that only students in condition 1 (no digital augmentation) did not demonstrate a significant increase in their understanding after manipulating the device. This finding suggests that the digital augmentation did have an impact on conceptual knowledge. The fact that students had the greatest gains in condition 2 (just digital augmentation and no other scaffolds) suggests that other scaffolds may not have been necessary to increase learning of these general concepts. However, the results found an increased higher order thinking from students in condition 4, suggesting that scaffolds might be necessary to reach more advanced learning. Both findings are consistent with other research. For example, Klopfer and Squire (2008) found that students were basically able to solve the simple problem, but required additional teacher supports to resolve more complex issues. Similarly, John and Lim (2007) found that although students learned from the medical training augmentations, the learning gains were enhanced when combined with pedagogical scaffolds. There are other plausible explanations for why differences in learning outcomes between the conditions were found in the conceptual knowledge survey. In research meetings, one researcher on the team discussed the idea that the scaffolds may have overly formalized what was supposed to be an informal experience, e.g., reading worksheets, following directions, and filling them in. This researcher was the main observation field note taker. She recalled that the feeling in the room was markedly different between the conditions. Students in condition 2 were generally more playful and experimenting on their own—much like students would normally behave on the museum floor. Students in conditions 3 and 4, however, were more serious and referred to the questions to dictate their next steps thereby missing the important concepts—much like traditional classroom instruction. Another explanation that may be related to this over-formalization is the fact that students in conditions 3 and 4 spent more time with the device and the post-intervention survey questions were the last thing standing in the way of running off to participate in other nearby museum exhibits such as the Sports Challenge. Thus, their rushed responses may not be an accurate indication of what they learned.

2) To what extent do knowledge-building scaffolds added to visualizations improve higher order inquiry skills?

As noted earlier, while there appeared to be no significant effect of the worksheet and knowledge-building scaffolds on increasing learning of concepts, students' abilities to generalize from their hands-on experiences, which we have interpreted as a higher order inquiry skill, showed differences across the conditions. The increasing trend in Figure 3 appears to be related to the number and kinds of scaffolds students were exposed to while manipulating the device. Students in condition 4 (with all scaffolds available) were better able to generalize their understanding than the other condition participants and the ability to generalize also appeared to relate to the presence of scaffolds. This finding is relevant to the field of informal science learning in that we provide some evidence that indicates that a modified knowledge-building approach may be useful in helping students learn science beyond simply manipulating and observing phenomena (NRC, 2009). Furthermore, through our investigation of a kind of learning design, we believe that this study responds to the call for more systematic research that identifies how learning science in informal science environments can impact the broader goals of STEM education (Rennie, 2003).

3) How can knowledge-building scaffolds, that are highly successful in sustained and continuous classroom activities, be adapted to support single-visit episodic and free choice engagement in museums?

Although the results for question #2 were encouraging, we were also interested in identifying which scaffolds were most helpful in terms of learning science in the museum. From the interview data analysis, it appears that the findings were mixed. Overall, students identified the ability to collaborate with each other as the most helpful scaffold, which is consistent with other knowledge-building studies (e.g., Yoon, 2008) and studies on collaborative learning (Dillenbourg & Schneider, 1995; O'Donnell & O'Kelly, 1994). One of the important design constraints that challenged the research team during construction of the condition activities was the notion of deep investigation over an extended period of time that is characteristic of successful knowledge-building classrooms (Scardamalia, 2002). We were concerned that the single-visit episodic nature of museum learning experiences would greatly inhibit the ability for students to knowledge build. To be clear, we are not claiming to have achieved knowledge building in this study. Rather, we are addressing the ability to use knowledge-building scaffolds to promote learning in informal environments. Those scaffolds were represented by the directions to collaborate, the collaboration itself, knowledge-building prompts, and the bank of other group responses intended to simulate the use of "rise aboves." From the interview data, there is evidence to suggest that there is some utility when students are familiar with the terms and processes. However, to ensure that all students receive the intended learning benefits of knowledge building, more time needs to be devoted to learning the terms and processes which would take longer than one activity during a single visit affords. This constraint in addition to our hunches about the over-formalization of informal activities and the increased higher order skills in the last condition has led the team to conclude that further investigation is warranted.

Conclusions and Future Research

Results of the study suggest that digital augmentations can help in conceptual development of science content. They also suggest that higher order thinking with respect to being able to generalize from the museum experience can be improved through the use of worksheets and knowledge-building scaffolds. While students unanimously said that collaborating in small groups was helpful and a majority of students said that the worksheets were helpful, the utility of the other scaffolds is inconclusive. Preparations for another study are underway with the following modifications. For C4, since it was not clear that students had actually read the directions, they will be read aloud. Rather than a suggestion, students will be required to consult the bank of previous answers before they construct their own. Students will also be instructed briefly about what the knowledge-building prompts mean and, finally, the post-intervention surveys will be administered the day after their museum visit rather than immediately after the activity. To address the issue of over-formalization, another condition will be added between conditions 3 and 4 in which we will post the worksheet questions but with the directions, knowledge-building prompts, and the bank of previous group answers added. In this condition, students will also work in small groups but will not be required to fill in the worksheet while they are interacting with the device.

References

Allen, S. (1997). Using scientific inquiry activities in exhibit explanations. *Science Education,81*(6), 715-734.

Bell, P., Lewenstein, B., Shouse, A., & Feder, M. (Eds.). (2009). *Learning Science in Informal Environments: People, Places and Pursuits.* Washington, DC: National Academy Press.

Borun, M. & Miller, M. (1980). *What's in a name? A study of the effectiveness of explanatory labels in a science museum.* Philadelphia: Franklin Institute.

Business Roundtable (2005). *Tapping America's Potential: The Education for Innovation Initiative.* [Web Resource] Available online. www.businessroundtable.org/publications.

Dillenbourg, P., & Schneider, D. (1995). Mediating the mechanisms which make collaborative learning sometimes effective. *International Journal of Educational Telecommunications, 1,* 131–146.

Domestic Policy Council (2006). *American Competitiveness Initiative*. Office of Science and Technology Policy: Washington, DC.

Dori, Y. & Belcher, J. (2005). Learning electromagnetism with visualization and active learning. In J. Gilbert (Ed.) *Models and Modeling in Science Education*, vol. 1 (pp. 187-216). Dordrecht, The Netherlands: Springer.

Falk, J., & Dierking, L. (1992). *The Museum Experience.* Washington, DC: Whalesback Books.

Hewitt, J. & Scardamalia, M. (1998). Design principles for distributed knowledge building processes. *Educational Psychology Review, 10*(1), 75-96.

Klopfer, E. & Squire, K. (2008). Environmental detectives - the development of an augmented reality platform for environmental simulations. *Education Technology Research and Development*, *56*, 203-228.

John, N. & Lim, I. (2007). Cybermedicine tools for communication and learning. *Journal of Visual Communication in Medicine*, *30*(1), 4-9.

National Science Foundation. (2006). America's pressing challenge: Building a stronger foundation, a companion to science and engineering indicators. Arlington, VA: Author.

O'Donnell, A. M., & O'Kelly, J. B. (1994). Learning from peers: Beyond the rhetoric of positive results. *Educational Psychology Review, 6,* 321–349.

Partnership for 21st Century Skills (2007). *The Intellectual and Policy Foundations of the 21st Century Skills Framework*. Tucson, AZ 1-24.

Rennie, L. J., Feher, E., Dierking, L.D. & Falk, J. H. (2003) Towards an agenda for advancing research on science learning in out-of-school settings. *Journal of Research in Science Teaching*, *40*(2), 112-120.

Scardamalia, M. & Bereiter, C. (1994). Computer support for knowledge-building communities. *Journal of the Learning Sciences, 3*(3), 265–283.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal Education in a Knowledge Society* (pp. 67-98). Chicago: Open Court.

Sengupta, P. & Wilensky, U. (2009). Learning electricity with NIELS: Thinking with electrons and thinking in levels. *International Journal of Computers, Math and Learning, 14*, 21-50.

Strauss, A. L. & Corbin, J. M. (1998). Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory (2nd edition), Sage, Newbury Park, CA.

Tasker, R. & Dalton, R. (2008). Visualizing the molecular world – design, evaluation, and use of animation. In J. Gilbert, M. Reiner, M. Nakhleh (Eds.), *Models and Modeling in Science Education*, vol. 3 (pp. 103-131). Dordrecht, The Netherlands: Springer.

U.S. Department of Education (2007). *Report of the Academic Competitiveness Council*. Washington, DC: Author.

Waite, T., Kirkley, J., Pendleton, R. & Turner, L. (2004). MUSEpad: Supporting information accessibility through mobile location-based technology, *TechTrends*, *49*(3), 76-82.

Yoon, S. A. (2008). An evolutionary approach to harnessing complex systems thinking in the science and technology classroom. *International Journal of Science Education*. *30*(1), 1-32.

Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant No. 0741659. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Beyond Small Groups: New Opportunities for Research in Computer-Supported *Collective* Learning

Yasmin B. Kafai, University of Pennsylvania, 3700 Walnut Street, Philadelphia PA 10194, kafai@upenn.edu
Kylie A. Peppler, Indiana University, 201 Rose Avenue, Bloomington, IN 47405, kpeppler@indiana.edu

Abstract: CSCL research has focused on understanding and designing collaborative learning in diverse settings and configurations with support of computers. Within this research, however, most efforts have concentrated on studying small group configurations and thus examined what we would like to call 'collaborative' learning (i.e., the abilities needed to participate and support collaborations of typically two to five people). Much less emphasis has been placed on studying massive communities and participation in large groups prominent in today's social networking sites and online gaming cultures that would shift the focus to 'collective' learning (i.e., the abilities needed to participate and support collaborations in massive groups). In this paper, we identify key dimensions of collective learning, present observations of online and local participation in one open-source Web 2.0 community with over 630,000 members, called Scratch (scratch.mit.edu), and outline a research agenda for computer-supported *collective* learning.

Introduction

Research in CSCL, by its very name, has focused on understanding various dimensions of group work such as productivity of different group arrangements (Engelmann & Hesse, 2010), design of scaffolds (van der Pol, Admiraal, & Simons, 2006), argumentation practices (Scheuer, Loll, Pinkwart, & McLaren, 2010) and interactions between online and offline collaborations (Birchfield & Megowan-Romanowicz, 2009). With some exceptions (e.g., Fields & Kafai, 2009; Guzdial & Rick, 2006), there is one assumption about collaboration underpinning many of these efforts, which is the idea that collaboration happens in small groups, often of dyads and triads, as they engage in computer-supported collaborative tasks. For the most part, the focus of this research has been to understand and develop what we will refer to as *collaborative* learning because it emphasizes the abilities to participate in small groups whether online or offline or in combinations thereof.

Recent developments, however, suggest new forms of collaborations are developing in online communities (e.g., Boyd, 2008; Shirky, 2008). One striking feature of these communities is their size and collaboration that can take place among hundreds, if not thousands, of members. Consider the millions of contributions to entries in the Wikipedia or to programs in Linux (Benkler, 2006), the interactions of members of fan fiction sites where thousands of writers create new stories and participants provide constructive feedback (Black, 2006), or the participation in guilds in multiplayer online role-playing game communities with millions of players (Gee, 2003). We have chosen to call this type of collaboration *collective* learning because it emphasizes the abilities to participate and perform in collectives and thus might be different from participation in small groups. Our concept of collective learning is inspired by recent research in different communities: the work of social scientist Pierre Levy on collective intelligence (1997) examining the potential of intellectual contributions from large groups; the work of gaming researcher Jane McGonigal (2008) observing self-organized coordination among players in collective gaming; and the work of media scholar, Henry Jenkins (2006) studying participatory culture in networked communities. Taken together, this body of work converges, helping us to recognize that large-scale communities can promote new forms of opportunities, as well as challenges, for learning together.

As increasing numbers of learning communities in K-12 and higher education move online, so too grows our need for understanding how to engage large groups of learners effectively in these networked communities. The starting point for our investigation on collective learning is a simple question: Does the size of the group matter? We realize that there will be no simple answer because massive communities such as social networking sites, virtual worlds, and multiplayer online role-playing games are each organized around different purposes and practices and thus participants have different incentives for joining and collaborating with each other (see also Hung, Lim, Chen, & Koh, 2008). We begin with a review of what research has identified as key features of collaborative learning in small and large groups. We then draw upon research and our knowledge of one such large-scale networked community, called Scratch (scratch.mit.edu), to describe and articulate the different dimensions of collective learning. In our discussion, we outline the emerging challenges and opportunities for research in computer supported *collective* learning.

Background

The work in CSCL draws upon hundreds, if not thousands, of research studies that have investigated various aspects of collaboration, including the nature of various group arrangements such as reciprocal teaching or jigsaw techniques, interactions with members of different gender, race, ability, and experience, and causes for success and failures of group work (O'Donnell, 2006; Webb & Palincsar, 1996). The research on computer supported collaborative learning, as documented in the *International Journal of Computer-Supported Collaborative Learning* (ijCSCL), has built on this substantial body of research and provides compelling evidence that collaboration can be arranged and even fostered in instrumental ways via computer support, albeit with a focus on what we would like to call *collaborative* learning – the ability to interact in and contribute to small groups. By contrast, the research on *collective* learning – the ability to interact in and contribute to large groups – is only now beginning to build such a body of work. We call this collective learning because groups are described by being in a particular space such as a virtual world or gaming community and often focused on tasks such as contributing to a collaborative repository.

To date, only a few studies in *ijCSCL* have tackled aspects of collective learning by designing wiki activities for university classes (Guzdial & Rick, 2006) or studying an afterschool gaming club participating in a tween virtual world with over 1.5 million players (Fields & Kafai, 2009). Some researchers have begun to conceptually map out issues of understanding collaboration and learning in Web 2.0 communities (Dohn, 2009) as well as to identify issues of trust building (Gerdes, 2010), while others have argued against adopting these models for school communities (Hung, Lim, Chen, & Koh, 2008). The size in number of participating members in these online communities is indeed impressive, especially considering that the possibility to interact with hundreds of thousands, if not millions, of others across geographical boundaries is unprecedented in history. This is particularly true for youth who might have connected at most with hundreds or thousands of others in their local communities and schools. Research is now showing that participation in these online communities can provide rich opportunities for learning to solve complex problems and for learning to collaborate with others in large groups (boyd, 2008; Gee, 2003; Ito et al., 2008). Yet it is unclear what it means to participate effectively in large-scale groups, to develop and foster a sense of community and belonging, and to design for collective learning interactions.

Closest to issues associated with collective learning is the work on Computer Supported Intentional Learning Environments (CSILE) (Scardamalia & Bereiter, 1991) and other studies following the knowledge forum tradition, since they examine how students' knowledge-sharing, knowledge-construction, knowledge-creation, and knowledge-assessment (e.g., Ares, 2008; Eddy, Chan & van Aalst, 2006; von Aalst, 2009) come into play. Most CSILE implementations have operated within a classroom environment, sometimes connecting students from other classes or previous years, and thus are far away from the size of social media communities today. Despite this, Scardamalia (2002; see also Zhang, Scardamalia, Reeve, & Messina, 2009) articulated early on the need for collective cognitive responsibility describing it as the "conditions in which responsibility for the success for a group is distributed across all the members rather than being concentrated on the leader" in addition for taking on responsibility "for knowing, what needs to be known, and for insuring that others know what needs to be known" (p. 2). Concepts such as collective cognitive responsibility indicate that interactions in these online communities could have different constraints and affordances due to their massive number of participants, unstructured and structured groups with concurrent asynchronous and synchronous nature of interactions, ongoing persistence of online life in absence of individual presence in addition to the nature of collaborative tasks and contributions. To better understand what collaborative vs. collective learning in a large-scale networked community looks like, we will first use an illustrative case study of Scratch community to examine some of its features in context before highlighting similarities and differences between these two paradigms of social interaction.

Computer-Supported Collective Learning in Networked Community of Scratch

We have chosen as a case study one open-source Web 2.0 application, called Scratch (scratch.mit.edu), to illustrate different dimensions of both collaborative and collective learning. Scratch is a media-rich programming language that allows youth to design, share, and remix software programs in form of games, stories, and animations. Scratch uses a familiar building block command structure (Resnick, et al., 2009), eliminating thorny debugging processes and the risk of syntax errors (see Figure 1a). Furthermore, programmed objects can be any imported two-dimensional graphic image, hand-drawn or downloaded from the Web, to further personalize each project. This makes it particularly amenable to an array of novice programmers wanting to build their own software and engage in the participatory culture. Since its public launch in May of 2007, the Scratch website (http://scratch.mit.edu) has been the primary means for users to share their work with one another (see Figure 1b). With over 1.2 million registered contributors and over 630,000 projects shared to date, the Scratch website is a vibrant online community with over 1,000 new projects being uploaded every day. It is possible to use Scratch as an individual programming tool or in traditional small groups formats (e.g., pair programming) but the website facilitates remixing and open-source like sharing of programs with others across time and space. In the following section, we review prior work on the emerging Scratch community and reframe this work with respect to our paper's central thesis: collaborative versus collective learning.

Figure 1. (a) Screenshot Scratch Interface and (b) Screenshot of Scratch Website.

We first present the typical activities that are organized collaboratively and collectively within the Scratch community. Each of these activity structures occur simultaneously within the community serving separate ends and potential contribute to separate types of learning experiences. In order to better articulate the differences between collaborative and collective learning, we have compiled a table (see Table 1) that highlights the key differences and similarities between collaborative and collective learning. It's important to note that our example, the Scratch community, affords both types of learning simultaneously, and while layered, each has its own distinctions.

Collaborative activities include mostly goal-oriented activities. Typically in the classroom, these would include small group work, team projects, discussion groups and so on. Outside the classroom, collaborative activities include playing board games, most platform style videogames with members of your family, sitting on committees, or playing in a rock band. In the Scratch online community, collaborative activities typically revolve around the production of particular types of Scratch projects by a small group of individuals that have commonly met one another in the online environment. One study focused on observations of a small collaborative group within the Scratch community called "Green Bear Group" that had formed spontaneously on the Scratch website by three children ages 8, 13, and 15 (Aragon et al., 2009). Over time, this group expanded and was later joined by over a dozen other members. They posted their projects on a gallery and members vote on which projects to further develop since each member brought different skills such as music, graphics or editing to the group. Based on an analysis of comments on the gallery as well as a survey, researchers observed found 19% of comments related directly to the job that needed to get done, 49% on socio-emotional aspects such as socializing and personal discussions, and 32% on contextual aspects such as arranging how to organize work, system administration and hardware issues. This suggests that collaborative groups do more than just getting the job done, they can also provide social and emotional support. Moreover, while bearing similarities to collaborative groups found in many assigned classrooms, what distinguishes these types of online collaboratives from regular groups are that they're self-organized and with flexible roles.

Collective activities, by contrast, include larger groups of individual with participation that is less goal-oriented and revolves more about sustained, enjoyed participation within the community over time. These types of activities although sparse in classroom settings, are common particularly in online affinity groups and in the arts. In the performing arts in particular, collective activities are common including African dance circles or orchestra performances. Wikipedia, Facebook, and Massively-Multiplayer Online Games are examples of collective activities where there may be some goal-oriented activities that drive participation but the community is driven by many individuals making modest contributions or participating peripherally and by a few intense participants. Within the Scratch community, this move toward membership in a large-scale community like Scratch can be a complex interplay between how young software designers develop personal agency with programming and gain status as experts amongst their peers. For instance, we followed two 12-year old participants, Lucetta and Matthew, as they learned the programming software Scratch and then joined scratch.mit.edu both in an after-school club and in a class as part of a four-month long ethnographic study (Kafai, Burke & Fields, 2010). We found that the web community furthered both Lucetta and Matthew's membership in a programming community in different ways. Lucetta friended other users, commented on projects, and uploaded her own projects, taking advantage of the social community on the site. This fit her cooperative social style, mixing with others while sharing an interest in Scratch. In contrast, Matthew embraced the potential of remixing, though there were other aspects of participation that he did not take up such as sharing his own project for validation and feedback from the community. However, Lucetta's initial resistance to migrating online and Matthew's own reluctance to upload his club project to the Scratch website, also suggest that establishing membership in a larger programming community is not as easily achieved as we had hoped. This suggests that navigating the collective community can accommodate an array of participation strategies and that there is a range of ways to participate in the collective and that it's probably necessary to learn about how to participate in collective activities in ways that might be initially uncomfortable. For example, what's lost or gained in Matthew and Lucetta's participation? What are the learning benefits of each of their participation strategies? How might we teach them to participate in collectives in new ways, identifying with new practices? Moreover, what's the potential role that schools can play in this landscape?

Table 1: Collaborative vs. collective learning.

	Collaborative Learning	Collective Learning
Group size is…	Small groups with less than 5 individuals, and more typically 2-5 individuals	Large groups of more than 5 individuals; can involve whole class instruction, frequently groups of 30-40 or possibly even hundreds or thousands of individuals that embrace a common set of practices and ways of participation
Typical Activities include…	Also includes mostly goal-oriented activities (i.e., a group assignment in a course). Small group work; Team projects; Discussion groups; Board games; Platform Video games; Committees; Rock bands	Affinity groups; Wikipedia; Dance Circles; Massively Multiplayer Online Games; Sports teams; Orchestras
Time is…	Often relatively short, task-based time periods that can be either synchronous or asynchronous and is usually organized by the group leader.	Often needs to happen over long periods of time with cyclical patterns or trends of involvement that can be further analyzed. Most often this is asynchronous but there may be spurts of synchronous large-scale activity. Less organization to the way in which time is organized in the collective. The web-master has the most immediate impact on the organization of time (e.g., calling for work on a certain theme) but the community can arrange happenings as well (e.g., flash mobs).
Leadership is…	In classroom communities, it is often centralized (e.g., a teacher directs activity).	In online communities, leadership is emergent, decentralized, and distributed among a large group of individuals; Collective leadership is also an inherent to the role that designers play in shaping the online community but also co-constructed in the community.
Division of Labor is…	Usually prescribed; roles of either group leader and participants. Participants usually take on the same role in the group with one leader emerging from the group. Absenteeism affects the group. There is a need for every member to take on a strong role in the group – if one or more members are reticent members, the group's goals can be compromised.	Members take on one role to fill a legitimate community need within the larger group. Absenteeism affects the collective less until it becomes widespread. Roles are un-prescribed and fluid – individuals can inhabit a range of roles but don't take on all roles. Newcomers can be peripheral participants in the activity without affecting the collective goals.
Knowledge is…	Distributed among a small group of individuals that commonly have some face-to-face interaction. Less institutional history (i.e., course assignments are repeated year after year in the course with little communication among students)	Collective knowledge, by contrast, by de facto is stored on the main site or with links to other off-site information (perhaps a how-to webpage) and is readily accessible to the community to have large-scale impact. Collective knowledge is also distributed among individuals, commonly separated in time and space. Knowledge usually has some institutional memory as old-timers remain in the group, and CSCL Communities typically create shared artifacts to share tips with other.

The entire Scratch community, however, is not made from simply small collaboratives and individuals participating in the collective activities fostered by the website. There are also collective groups (i.e., large affinity groups) that work to produce work of a similar interest over an unspecified period of time. For example, and within the online community, similar themes emerge and spread quickly among members with similar interests, including projects around the Japanese manga character, Naruto, projects inspired by Logo Art, and other popular videogames. We have also observed the emergence of large-scale collectives in local contexts around the use of Scratch. For example, a sub-community of 30-40 Scratch programmers in South LA began making Scratch projects based on Low Riders—highly personalized cars often characterized by having low suspensions and original paint and hubcap designs (Peppler & Kafai, 2007). This local affinity group worked to produce a number of Low Rider Scratch projects over the course of a year. While working almost entirely offline, this community looks similar to affinity groups that have emerged within the online community as well.

Moreover, new practices emerge at the collective level that are simply not possible in small collaborative groups. For example, a particular form of participation within the Scratch is called "remixing", taking existing Scratch projects and changing them before uploading them back to the website. It has been argued that remixing is a key practice in today's networked culture in support of our knowledge production. Crediting ownership consists of referencing the intellectual origins of "text" used in media productions. For instance, remixing Scratch projects (i.e., taking an existing project and modifying code or graphics) is a common practice in the larger Scratch community; in fact, over 40% of all projects posted on the web site are remixes of existing Scratch projects (Senivirate & Monroy-Hernández, 2010) and this number has been steadily increasing since its launch. The culture of remixing in Scratch itself is not without fraught as positive and negative comments are prevalent. In an after-school club, Scratch programmers ages 10-12 years were adamant that their fellow programmers credited the origins of programs that they had remixed and posted online. While Scratch programmers initially were concerned about others taking their programs, they also came to understand the remixes as a form of recognition that represented attention they received from others (Kafai, Burke, & Fields, 2010).

In addition to the differences in group sizes and types of activities that collaborative and collective groups tend to engage in, there are several additional distinguishing features about collaborative learning versus collective learning, which we summarize here in the ways that collectives organize time, community leadership, the division of labor, and knowledge, which are further described below and draw upon the examples mentioned above. In collaborative groups, time is often organized in relatively short intervals in a typically synchronous but sometimes asynchronous fashion. Additionally, time is usually organized by the group's leader, which makes goal-oriented deadlines. For example, a project leader might organize a time for the group to meet or a deadline for materials to be received as in the case of the Green Bear Group. By contrast, time within a collective is more open and with cyclical patterns or windows/ opportunities for involvement at an upcoming event or happening. Most often though time in collectives is asynchronous and not deadline oriented. For example, online communities typically have round the clock involvement of its members, so that no matter what time you login there are always large groups of individuals in the environment. Time in the collective also demonstrates patterns of activity, which allow for trends to emerge over time with little intentional coordination from its members. While in the Scratch community there is no unified effort to post projects on a certain day or time around a specified theme, there is nevertheless certain trends or memes that spread quickly within the community similar to the Low Rider creators mentioned above. For example, there was a sudden posting and remixing of Mother's day projects in May 2010.

Leadership within collaborative groups usually falls upon a single individual or small group of individuals. While some aspects of this type of top-down leadership can still be seen in collectives, leadership is also distributed among its members, emergent, and decentralized. For example, the online community was tweaked by its designers to highlight "Top Remixed Projects" to promote more remixing within the community. After doing so, the number of projects claiming to have remixed another project went up substantially as well as the number of projects that were created as tools for others to build on and remix went up substantially (as is the case with projects like side-scrolling game engines). Individuals also began to make new types of projects that could be remixed and to advertise and compete for these slots on the homepage in ways that were unexpected by the webmasters. Similarly, prior studies in MMOs have demonstrated this type of emergent leadership and use of tools in unexpected ways that the designers hadn't intended.

Accordingly, the division of labor is quite different within these two types of groups. Within a collaborative group, members all take on particular identities within the group as the programmer, the designer, musician, and editor with a distinct contribution to the group's production goals. These roles are usually prescribed by the teacher or group leader and somewhat stagnant over the course of the project. Absenteeism also highly affects the group. There is a need for every member to take on a strong role in the group because if one or more members are reticent members, the group's goals can be compromised. Within the collective, members take on a role to fill a perceived need within the community. Absenteeism also affects the collective less until it becomes widespread as there are multiple individuals with similar expertise available to fill in and knowledge is dispersed amongst its members. Roles are also un-prescribed and fluid. Individuals can inhabit a range of roles but don't necessarily take on all roles and newcomers can be peripheral participants in the activity without affecting the collective goals.

Lastly, knowledge is public and well distributed among members of the Scratch community. This is most visible when you look at the community as a whole and seeing it growing and changing in many ways over time. One might wonder whether expertise grows over time in these informal communities, and, indeed it does. Based on data from the South LA field site of Scratch users, pre- and post-test analyses revealed that youth learned about the big ideas of computer programming as well as visual/multimodal media arts production over the course of two years in absence of direct instruction (Maloney, Peppler, Kafai, Resnick, & Rusk, 2008). These analyses were consistent for individual learning as well as community learning (i.e., new members were being apprenticed into the community to produce more sophisticated work over the course of the study). This body of research demonstrates that the collective can generate competence in a domain in the absence of direct instructions and what is traditionally thought to be subject area expertise. In fact, the corpus of projects growing on the Scratch website at exponential rates reflects a similar growing understanding made more accessible by the fact that any project can be downloaded, analyzed and easily remixed to enable new generations of members to achieve tasks like creating side-scrolling gaming engines more quickly and easily than the ones prior.

Discussion

In this paper we examined different dimensions of participation in massive networked communities. We return to our initial question: does the size of group matter? As expected, there is no simple answer to this question. Collective learning is really about the development of a new community of practice over time (Lave and Wenger, 1991). This is particularly true for a community like the Scratch online community, which has only emerged in the past few years. As a result, specific notions of what it means to share work and start new projects evolve in this community. But we also do know that practices around sharing information and collaborating have evolved differently than in other programming communities such as StarLogo or Alice where remixing pop culture texts like Naruto or Manga is not core to participation in the community. While some aspects, for instance the voluntary forms of peer production and contribution for the common good, suggest that new forms of collaborations are emerging, others such as the formation of small groups and issues of intellectual ownership suggest very much a continuation of previous collaborative practices. We have organized our discussion around the following four main themes: groups, designs, ethics, and research in collective learning environments. This list is by no means comprehensive but it serves as a promising starting point for thinking about new research opportunities in computer-supported collective learning.

Our first theme focuses on the group as the main unit of analysis. This has been a focal point in collaboration research, and to some extent, continues to be a focal point for research in massive communities as we observed in the formation of smaller groups in Scratch online and offline communities. It is clear that we need more research to understand how such groups are being formed, who decides to join, and how these groups continue working together, and also when and why they fail in their efforts. We also need to better understand differences in group formation and how this relates to the nature of artifacts produced. For instance, in many gaming communities players organize in guilds to engage in quests. These guilds are highly structured organizations where players assume different roles and contribute to the success of the mission (e.g., Gee, 2003). Contrast this with the free formation of the "Green Bear Group" in Scratch that set out to design games. Here members contributed based on their expertise but roles could change based on needs. The traditional collaboration research (Cohen, 1994) as well as CSCL research provides us with little information about the dynamics of unstructured group collaborations as prevalent in many massive communities. We can say that one aspect of collective learning is to assume multiple roles; these roles are not prescribed though members are valued and recognized in the community for their particular abilities. Such changes in participation are also part of communities of practice and often assumed for successful collaboration in small groups. Perhaps it is the prescriptive nature of assigned roles in small group collaboration that make it difficult for members to adopt the fluidity needed to function and perform and larger collectives.

Technical designs are equally important in studying collective learning since massive networked communities are artifacts, meaning that structures and features are designed by programmers and modified through community use. There are multiple feedback mechanisms and documentation notes that can be integrated and made accessible to participants. For instance, we do not understand which features lead members to contribute productively to large efforts and which ones might hinder such contributions. While some of these, such as recognizing heritage in remixed projects are social issues, they also involve technical solutions. For instance, in Scratch, the system automatically keeps track and makes visible such lineage. On the other hand, we have participants who are playing a growing role in content production, in the case of many virtual worlds such as Second Life, even are the main content producers for the community. How much of this control is ceded to participants is something that needs to be considered in the setup of these community as well as the technical prerequisites of lay designers themselves, in particular when we talk about younger participants.

We also need to consider issues about ethics that invariably come up in the context of collective production, sharing and commenting and are particularly relevant in the context of schooling that still favors individual recognition. What does it mean when members contribute to a larger project, how is this recognized? What about remixes that take up existing projects and modify them – a practice very common in networked communities? In a recent ijCSCL article, researcher Gerdes (2008) raised the issue of developing trust that surely plays a growing role in massive communities where participants do not know each other and perhaps never meet face to face.

Finally we need to expand our repertoire of research methods in describing and analyzing collective learning. Studies of gaming and social network communities (Boelstorff, Taylor, Nardi & Pearce, in press; Hine, 2000; Williams, Yee, & Caplan, 2008) reveal an unhealthy split in either quantitative or qualitative research approaches. For instance, survey methods and statistical data mining seem to drive many efforts in coming to grips what engages members in these communities. On the other end, we have ethnographies of single massive community (see Boelstorff, 2008; Taylor, 2006) that inform us with a fine-grained detail of cultural practices and activities. Of course, others have complained about this dichotomy and argued for a mixed methods approach (Williams, 2005). In our view it is not just about juxtaposing data sources and analytical methods but also about developing ways that integrate both in a productive manner. As a case in point, we have suggested and employed connected ethnographies that make use of the data mining and reduction in large data sets to identify particular participants based on their contribution profiles and to cross reference and develop these through in-depth ethnographies (Kafai & Fields, in press; see also Reimann (2009). Such analyses leverage the explanatory potential of each method and allow us to contextualize cases within larger community trends.

All of these features come into prominence when we think about the design of massive intentional learning or knowledge building communities (Scardamalia & Bereiter, 1991) that design for collective learning. In the case of Scratch, we have started collaborative design challenges in the online community that invite members to work together on programming projects. Our efforts focus on understanding how groups and collectives form and interact across global and local Scratch communities and how to better design and support computer-supported collaborative and collective learning. Our observations suggest that it might be worthwhile to think about the second C in CSCL not just as collaborative but also as collective learning. It's not just a simple matter of involving larger numbers of participants but also of considering the nature of activities, the roles that participants will assume, and the performances or artifacts that present the culmination of efforts to effectively design and study collective learning environments. We suggest that CSCL researchers step outside the boundaries of small groups and begin to consider alternative arrangements. Both collaborative and collective learning technologies, activities, and environments that we design are greatly influenced by how we think about learning as a socially situated activity. Size does matter – we just need to know how and when to engage in collective vs. collaborative learning.

References

Aragon, C., Poon, S., Monroy-Hernandez, A., & Aragon, D. (2009). A tale of two online communities: Fostering Collaboration and Creativity in Scientists and Children. In *Proceedings of the ACM Creativity and Cognition* Conference, Berkeley, CA.

Ares, N. (2008) Cultural practices in networked classroom learning environments. *International Journal of Computer Supported Collaborative Learning, 3*(3), 301-326.

Benkler, Y. (2006). *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. New Haven and London: Yale University Press

Birchfield, D. & Megowan-Romanowicz, C. (2009) Earth science learning in SMALLab: A design experiment for mixed reality. *International Journal of Computer Supported Collaborative Learning, 4* (4), 403-421.

Black, R.W. (2008). *Adolescents and Online Fan Fiction*. New York: Peter Lang.

Boellstorff, T. (2008). *Coming of age in second life: An anthropologist explores the virtually human.* Princeton, NJ: Princeton University Press.

Boellstorff, T., Nardi, B., Pearce, C. & Taylor, TL (in press) (Eds.). *A Handbook of Ethnographic Methods for Virtual Worlds*. Princeton, NJ: Princeton University Press.

boyd, d. (2008). Why Youth (Heart) Social Network Sites: The role of Networked Publics in Teenage Social Life. In D. Buckingham (Ed.), *Youth, Identity and Digital Media* (pp. 119-142). Cambridge, MA: MIT Press.

Brown, A., & Campione, J. (1994). Guided discovery in a community of learners. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 229-290). Cambridge, MA: MIT Press.

Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research, 64*, 1-35.

Dohn, N. (2009). Web 2.0: Inherent tensions and evident challenges for education. *International Journal of Computer Supported Collaborative Learning, 4*(3), 343-363.

Eddy, Y.C.L., Chan, C., & van Aalst, J. (2006). Students assessing their own collaborative knowledge building. *International Journal of Computer Supported Collaborative Learning, 1*(1), 57-87.

Engelmann, T. & Hesse, F.W. (2010) How digital concept maps about the collaborators' knowledge and information influence computer-supported collaborative problem solving. *ijcscl* 5 (3), pp. 299-320.

Fields, D. A. & Kafai, Y. B. (2009). "U wanna go to the moon?" A connective ethnography of peer knowledge sharing and diffusion in a tween virtual world. *International Journal of Computer-Supported Collaborative Learning, 4*(1), 47–68.

Gerdes, A. (2010) Revealing preconditions for trustful collaboration in CSCL. *International Journal of Computer Supported Collaborative Learning,* 5 (3), 345-353.

Hill, M. B., Monroy-Hernández, A., & Olson, K. (2010). Responses to Remixing on a social media sharing website. *Fourth International AAAI Conference on Weblogs and Social Media*.

Hine, C. (2000). *Virtual ethnography*. London: Sage Publications.

Hung, D., Lim, D., Chen, V. & Koh, T. S. (2008) Leveraging Online Communities in Fostering Adaptive Schools. *International Journal of Computer Supported Collaborative Learning,3*(4), 373-386.

Jenkins, H., Clinton, K., Purushotma, R., Robison, A., & Weigel, M. (2006). Confronting the challenges of participation culture: Media education for the 21st century. White paper. Chicago, IL: The John D. and Catherine T. MacArthur Foundation.

Kafai, Y. B. & Fields, D. A. (forthcoming). Understanding Player Participation and Practices in Virtual Worlds: A Proposal for Qualitative Analyses of Log File Data. In D. Thomas (Ed.), *Research methods in virtual worlds*. Cambridge, MA: The MIT Press.

Knobel, M. & Lankshear, C. (2009) Wikis, digital literacies, and professional growth *Journal of Adolescent & Adult Literacy*, *52*(7), 631-634.

Leadbeater, C. & Miller, P. (2004). The Pro-Am Revolution: How enthusiasts are changing are economy and society. London: Demos.
Levy, P. (1997). *Collective intelligence.* Cambridge, MA: Helix Books.
McGonigal, J. (2008). 'Why I love bees: A case study in collective intelligence gaming' in K. Salen (Ed.), *The ecology of games: Connecting youth, games, and learning* (pp. 199-228). Cambridge, MA: MIT Press,
Maloney, J., Peppler, K., Kafai,Y., Resnick, M., & Rusk, N. (2008, March). *Programming by choice: Urban youth learning programming with Scratch*. Paper presented at the SIGCSE 2008 Conference, Portland, OR.
Monroy-Hernandez, A. & Resnick, M. (2008). Empowering kids to create and share programmable media. *Interactions*, *15*(2), 50-53.
O'Donnell, A. M. (2006). The role of peers and group learning. In P. Alexander & P. Winne (Eds.), *Handbook of educational psychology*, 2nd Edition. Mahwah, NJ: Lawrence Erlbaum.
Osterman, K. F. (2000). Students' Need for Belonging in the School Community. *Review of Educational Research, 70*(3), 323-367.
Peppler, K. A. & Kafai, Y. B. (2007). From SuperGoo to Scratch: Exploring creative digital media production in informal learning. *Learning, Media, and Technology*, *32*(2), 149–166.
Reimann, P. (2009) Time is precious: Variable- and event-centered approaches to process analysis in CSCL research. *International Journal of Computer Supported Collaborative Learning, 4*(3), pp. 239-257.
Resnick, M., Maloney, J., Author 2 Rusk, N., Eastmond, E., Brennan, K., Millner, A., Rosenbaum, E., Silver, J., Silverman, B., & Author. (2009). Scratch: Programming for all. *Communication of the ACM, 52*(11), 60-67.
Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Eds.), *Liberal education in a knowledge society* (pp. 67–98). Chicago: Open Court.
Scheuer, O., Loll, F., Pinkwart, N., & McLaren, B. M. (2010) Computer-supported argumentation: A review of the state of the art. *International Journal of Computer Supported Collaborative Learning, 5* (1), pp. 43-102.
Senivirate, O. & Monroy-Hernández, A. (2010). Remix culture on the web: A survey of content reuse on different user-generated content web sites. *Web Science Conference*. Raleigh, NC.
Shirky, C. (2008). *Here Comes Everybody: The Power of Organizing Without Organizations.* New York, NY: Penguin Press.
van Aalst, J. (2009) Distinguishing knowledge-sharing, knowledge-construction, and knowledge-creation discourses. *International Journal of Computer Supported Collaborative Learning, 4*(3), pp. 259-287
Webb, N. & Palincsar, A. (1996). Collaborative learning. In D. Berliner (Ed.), *Handbook of Educational Psychology* (pp. 345-413), New York: Macmillan.
Williams, D. (2005). Bridging the methodological divide in game research. *Simulation & Gaming, 36*(4), 1-17.
Williams, D., Yee, N. & Caplan, S. (2008). Who plays, how much, and why? Debunking the stereotypical gamer profile. *Journal of Computer Mediated Communication*, *13*(4), 993-1028.
Zhang, J., Scardamalia, M., Reeve, R., & Messina, R. (2009). Designs for Collective Cognitive Responsibility in Knowledge-Building Communities. *The Journal of the Learning Sciences, 18(*1), 7-44.

Acknowledgments

The writing of this paper was supported by a grant from the National Science Foundation (NSF-CDI-1027736) to the first author and a grant from the John D. and Catherine T. MacArthur Foundation's Digital Media and Learning Initiative to the second author. The views expressed are those of the authors and do not necessarily represent the views of the Foundations, University of Pennsylvania or Indiana University.

Visualisations for Longitudinal Participation, Contribution and Progress of a Collaborative Task at the Tabletop

Roberto Martinez, Judy Kay, Kalina Yacef, School of Information Technologies, University of Sydney, Sydney, NSW 2006, Australia
Email: roberto@it.usyd.edu.au, judy@it.usyd.edu.au, kalina@it.usyd.edu.au

Abstract: One of the challenges for facilitators in collaborative work is that they typically see only the final product of a groups' interactive work. This is a problem as it means that the role of each individual may be hard to determine. This paper proposes a set of visualisations which aim to give teachers insights into longitudinal participation of each group member, an indication of the extent of each learner's contribution and the building process of the group product in terms of overall activity towards a good solution. We exploit the affordances of tabletops to capture the data in order to infer these visualisations. We evaluate these by assessing whether facilitators could answer key questions about aspects of groups. Key contributions of the paper are the design of new visualisations, results of their evaluation and the implementation of a tabletop concept mapping application which was carefully designed to both support collaboration and capture of the history of the collaborative process.

Introduction

One of the challenges for teachers and facilitators in collaborative learning contexts is that they typically see only the final product of a group's interactive work (Race, 2001). This is a problem as it means that the group work itself and the relative contribution of each individual to the group product may be hard to determine. This problem can dissuade teachers from using group work. Yet, there is acknowledged value in collaborative knowledge creation in which learners create artefacts, such as tools, symbols, concepts, and texts; and make use of processes like grounding and negotiation (Paavola & Hakkarainen, 2005). Such artefacts can capture the common understanding of subject matter and innovative ideas of the group. One method to facilitate discussion of a content area is through the construction of concept maps (Suthers, 2008). We have chosen this well studied technique for our work as it has proven useful in education for externalising knowledge. Concept maps can be considered as external reflectors of internal learner's knowledge (Tergan, 2005).

Interactive tabletops provide a new way to support collaborative tasks because they permit face-to-face interactions between individuals and, at the same time, present information to the group through a shared device. However, they also introduce new challenges, both at the technical and social levels. Our research explores the affordances and challenges of supporting groups to externalise their knowledge of a domain by providing a tabletop interface and making the mapping process visible to learners' teachers.

We present the design and evaluation of a set of visualisations that reflect the activity of a small group of users building an artefact, such as a concept map, using a multi-touch tabletop. These visualisations provide a mirror of the learners' actions (touch/verbal participation radar), an indication of the extent of each learner's contribution (contribution chart) and the evolution of the artefact through its building process (evolution diagram) in terms of complexity of the group product. We aim to determine the visualisation features that give useful information about the collaboration of the group at the tabletop. These visualisations aim to summarise and make visible the actions of the group members to help facilitators detect problems in group interactions.

We evaluated the visualisations by assessing whether teachers could answer key questions about *participation*, *contribution* and the *process* of the collaborative activity of the group. Our key contributions are the design of these new visualisations, the implementation of a tabletop application for concept mapping and an evaluation which indicates that teachers can answer 4 of the 5 key questions related to longitudinal equity and quantity of participation, contribution of group members and collaboration.

The remaining of this paper is structured as follows. In the next section, we present our tabletop application for concept mapping and the design of our visualisations. Then, we describe the evaluation. We conclude with reflections on the visualisations and our research agenda.

Related Work

Several researchers have explored how to exploit the potential of digital tabletops in educational contexts. Morgan and Butler (2009) proposed settings that encourage learning with high levels of collaboration at multi-touch displays, grounded in theories of *Situated Cognition*. They designed systems for storyboarding, group concept mapping and building phonemic awareness between dyads. Busine et. al. (2007) described the use of interactive tabletops to support group creativity through the construction of mind-maps. Additionally, Sugimoto et al. (2004) showed how tabletops, used in conjunction with personal devices, can support discussion and negotiation in face-to-face collaboration with learners manipulating private and shared objects.

Research in tangible tabletops has also explored their affordances for detecting interference in collaboration (Falcao & Price, 2009), analysing division of labour (Jermann et al., 2009) and to aid mind-mapping (Oppl & Stary, 2009). However, a recent study failed to show significant learning benefits at a tabletop compared with working on a single desktop computer (Do-Lenh, Kaplan, & Dillenbourg, 2009). In this study, those learners using the multi-touch tabletop tended to work in parallel and had difficulties drawing all the pieces together at the end of the trial. This finding highlights the need for an effective support at the tabletops.

There has been considerable work exploring the importance of group visualisations to externalise the activity of groups and, in many cases, to reveal relationships between observable patterns and the quality of the group work. Erickson et al. (1999) created the *social proxy*, a visualisation of chat sessions of a group. This resulted in improved collaboration and better support for people to learn how to collaborate. In the same way, *sociograms* have been extensively used in the CSCL field to visualise learner interactions and they have also been applied for representing the lines of communication within social networks (Sundararajan, 2010). Donath (2002) went a step further, showing the *nature of participation* in a visualisation of group activities. In (Kay, Maisonneuve, Yacef, & Reimann, 2006) a set of visual representations of long term activity of groups was designed, based on a model of small group teamwork. These present salient aspects of group activity, including the level of participation, interaction between members and leadership. The *Narcissus* project introduced a visualisation of group activity which enables a user to navigate through to see the detailed evidence that contributed to each part of the group work (Upton & Kay, 2009). In addition, Janssen et al. (2007) explored the effects of visualisation of participation in groups of learners. They found that visual representations of activity mirrored to the group can be useful for encouraging coordination and regulation of the members.

The Collaborative Learning Environment

We now describe the context of our work to design and then evaluate visualisations of group activity in a collaborative concept mapping system. We used the tabletop system called Cmate (Martinez, Kay, & Yacef, 2010). This was designed to be used in two stages: first, an individual stage, where each group member creates their own concept map; and a group stage, where they come together to create a collaborative map. This approach was also used by Engelmann & Hesse (2010) with the difference that we provided a shared tabletop for the group phase. For the individual stage, participants used CmapTools (Novak & Cañas, 2008), a desktop-based concept mapping application. These individual concept maps are used to extract the personal vocabulary of concepts and links, and make them available to the learners in the group stage at the tabletop. Importantly, this stage enables each individual to focus on the cognitively demanding task of externalising their knowledge, by drawing their individual maps in private. Then, they come to the tabletop to compare and discuss their different perspectives of the subject matter and create a collaborative concept map (left of Figure 1). In this stage, participants can use the concepts and links they previously included in their individual artefacts and relate them with other participants' pieces of map to build a new mutually accepted map.

We designed Cmate to maintain the collaboratively created group map as well as one layer per user, showing their individual map contributions. The basic actions that learners can perform are: *create a concept*, drawing a circular outline with the finger; *create directed link*, drawing a line in between two nodes; *move concept/link,* touching the node and dragging it with the finger; *delete a node*, moving the node to off-screen; and *edit a node word*, double tapping a node and modifying the word through a virtual keyboard.

The tabletop used in the studies had a 46-inch LCD touch screen with a display resolution of 1920x1080 pixels, offering enough space for up to four participants. The tabletop hardware can detect multiple touches at a time, but cannot recognise which user is providing an input. For our study, every touch on the tabletop was logged and the entire sessions were videotaped using a fixed camera above the tabletop and another camera at one side. Sound was recorded with individual headphones worn by each participant.

In order to aid facilitators, it is crucial to know who is doing what and what has been done on the tabletop. Morgan and Butler (2009) proposed four approaches to track the contribution made by each person at

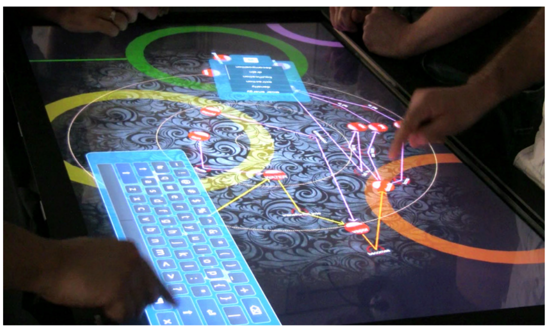


Figure 1. Tabletop concept mapping. Cmate in action with 4 people, 2 at each long edge of the table (*left*). Personal spaces for tracking individual participation on the tabletop (*right*).

the tabletop: by defining fixed workspaces where each person should work; by claiming "ownership" on the objects; establishing a production line, in which each object is sent to a task bar of each person at least once to work on it; or by defining roles, in which each person is allowed to do just some specific actions. We created a different approach to track individual participation. We provided a *moveable circular personal area* on the tabletop for each participant. Users can *initiate* all actions just inbounds of these areas. To perform actions on the concept map, participants move their personal area above the target element and then perform the actions. This technique increases the load of touches on the tabletop but gives personalised concept/linking word lists for creating nodes, orients the elements towards the user and also supports tracking of all the touches that each participant does (*right* of Figure 1). For the position of learners around the tabletop, it has been shown that this has an effect on the division of labour spontaneously adopted by the learners (Jermann, et al., 2009). Consequently, we had participants in pairs along both long sides of the rectangular tabletop. This gave each participant equal opportunity to participate, access the resources, and perform the full range of actions. Where the group had three participants, we had to settle for a slightly less ideal spatial disposition.

We ran the study with 10 participants in 3 groups, each of 3 or 4 participants. All groups were asked to build concept maps individually at a desktop and then a group map at the tabletop, on the topic: *how does the water cycle work?* First, participants were introduced on the concept mapping technique. Then they read a two page text on the water cycle and were asked to draw an individual concept map. Participants could use any terms, any number of concepts in their maps and were not restricted to follow any hierarchical arrangement of concepts. However, we provided a list of suggested words, extracted from the instructional text. After completing their concept maps the learners were asked to generate a group solution concept map at the tabletop.

Design of Visualisations

We now describe the four visualisations we designed: the Touch/verbal Participation Radar, the Map Contribution Chart and the Map Evolution Diagram.

Touch/Verbal Participation Radar

This first visualisation was strongly influenced by the circular social proxies of Erickson (1999). We also drew on previous work in collaborative learning, with a focus on the learning impact of the equity of oral participation and decision making (Bachour, Kaplan, & Dillenbourg, 2010). In addition, grounding on the procedure to study levels of participation at the tabletop proposed by Harris et. al. (2009), we chose to focus on two dimensions: the physical events on the tabletop, measured in terms of the quantity of touches; and time of verbal participation, measured in seconds.

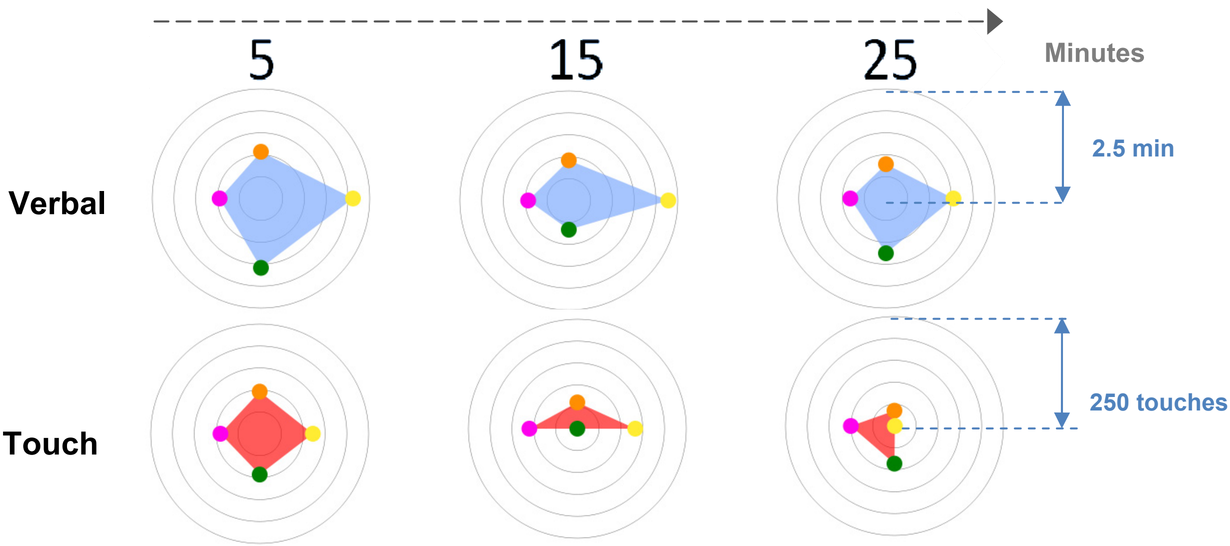


Figure 2. Participation radar. (*left*) First five minutes. (*center*) Between minute 10 and 15 (*right*) Between minute 20 and 25.

As a result, we designed a pair of radars: the radar of verbal participation (top of Figure 2, blue shaded radars) and the radar of touch participation (bottom of Figure 2, red shaded radars).The time window for each visualisation covers the previous 5 minutes of activity. So for example at time 15, the radars show the number of events between minutes 10 and 15. Each coloured round marker corresponds to one learner at their circular personal space: orange, yellow, green and purple for participants 1, 2, 3 and 4 respectively. The position of these markers indicates the level of participation; the closer the marker is to the centre, the less active they were in the last five minutes.

The shape of the radars depicts the symmetry of activity, an important aspect of collaboration (Dillenbourg, 1998). For example if there are 4 learners, a perfect symmetric square indicates that the number of touches or the talking time are the same for each learner. In the radars shown at the right of Figure 2 (at time 25), we can observe that the participant corresponding to the yellow marker (at 3 o'clock) did not touch the tabletop at all but did most of the talking. This could possibly be a clue that the participant was influencing the actions of others by talking but, without further information, it could equally means that he/she was engaged in a conversation that had nothing to do with the task.

Contribution Chart

Concept mapping has the potential to drive collaboration because it provides an externalisation of each learner's different perspective (Tergan, 2005). For this reason, we propose the contribution chart, which shows the proportion of actions from each team member that resulted in a change in the collaborative artefact. The actions which add substantial knowledge to the map are the creation, editing and deletion of nodes. The map contribution chart gives an overview visualisation of the proportion of these actions that each participant performed. Figure 3 shows three of these charts for the same group, at minutes 5, 10 and 20, and also shows the total number of concepts and links created in the group artefact. The design of this visualisation aims to show if any of the participant's perspectives is leading the construction of the concept map or if they are equitably contributing to it. This visualisation complements the radar of participation by indicating the amount of activity that has indeed made a substantial impact on the group artefact. For instance, the first chart in Figure 3 (at minute 5) shows that one person (in pink at the top) contributed about half the concept map's elements, two people (green and yellow) contributed a quarter, and one (red, at the right) much less than the others. At time 10, however, we can observe that the other three participants increased their contribution greatly (especially yellow and red), and later on (at time 20), the same participant (pink) contributed the most to the artefact again. The relative dimension of the contribution between two given moments is indicated by the overall number of links and concepts and also symbolised through the relative size of the charts.

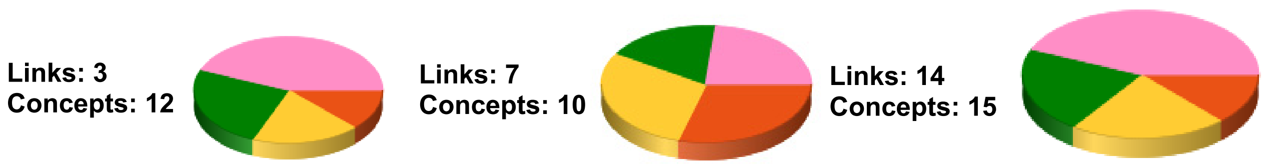


Figure 3. Map contribution chart. (*left*) After 5 minutes. (*centre*) After 10 minutes. (*right*) At after 20 minutes.

Evolution Diagram

The third visualisation is the Evolution Diagram (Figure 4). This shows the key temporal events in the group concept map. The vertical axis shows the number of propositions added to the group concept map and the horizontal axis represents time measured in minutes. The graph has two sparklines: i) the upper blue line includes coloured circles indicating addition of each proposition by a given participant; ii) the lower red line shows the number of propositions that match those in the master concept map (created by the teacher). So, the upper line shows the total number of propositions and the lower line shows how many of these propositions match the expert map. To calculate the distance between the group map and the master map, we use an automatic open-ended concept map scoring technique based on Pathfinder associative networks (Taricani & Clariana, 2006).

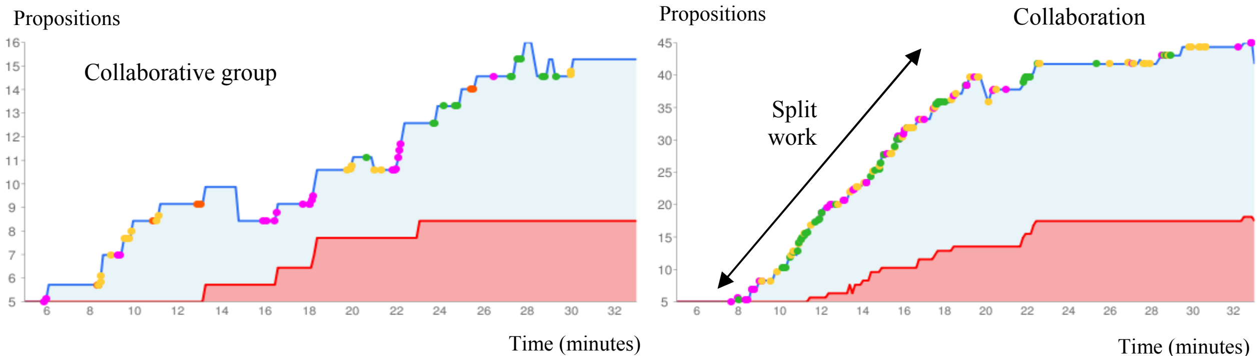


Figure 4. Map evolution diagram. (*left*) A group of four learners working collaboratively most of time. (*right*) A group of three learners who divided the work during the first 20 minutes.

In the upper line, the coloured markers represent the user who added the proposition(s). In Figure 4 (*left*), the visualisation shows that the group has worked for more than thirty minutes and has created 16 propositions. Observing the purple markers (User 4) we see that this user created more propositions that were present in the master map, as the purple markers coincide with signs of progress of the group map towards the master map. In contrast, the map evolution diagram on the right shows that, up to Minute 20, this group added many links to the map but these actions did not result in any matching with the expert map, suggesting that users were working independently. Note that, in this visualisation, the similarity with a master concept map is not used to score the group map but just as an indicator of the quantity of the group's propositions that match those in the teacher's perspective.

Evaluation

The aim of the evaluation is to ask a number of facilitators to give answer to a set of questions regarding the equity of participation, quantity of participation, collaboration, equity of intellectual contribution of the members of the group (Dillenbourg, 1998; Stahl, 2006) and the process of the concept map construction. We expect that the support offered by our visualisations should make visible some facets of the collaborative process to facilitators and hence lead to improve the feedback they can offer to the group. Specifically, we aimed to evaluate five hypotheses: The set of visualisations provide useful information about:

- (H1) the *equity in the roles and participation* of group members;
- (H2) the *amount of participation* of the group members;
- (H3) the group in terms of *collaboration*;
- (H4) the *equity of intellectual contribution* of group members;
- (H5) the *creation process* of the map in terms of the relative contributions of group members and the effectiveness of the group work.

To assess these hypotheses, we ran evaluation sessions with five different skilled facilitators with experience in working with groups. We provided them with a set of the visualisations generated from the tabletop sessions of two groups: A and B. These visualisations included: snapshots of the Radars of Participation and Contribution Chart corresponding to minutes 5, 10, 15, 20, 25 and 30, the Evolution Diagram and the final group concept map. We also provided the final maps, from both the individual and collaborative stages. Then, we invited the facilitators to respond a set of questions regarding the equity of participation, quantity of participation, the collaboration, the equity of intellectual contribution of the group members and the usefulness of the visualisations to depict the creation map process in terms of contribution and group work. Each question corresponded to one of the hypotheses posed above, and they were answered on a 7-point Likert scale in which 1 represents strong disagreement and 7 strong agreement. We allowed them to mark a question as "unanswered" if the visualisations did not give enough support for taking a decision. We asked the facilitators to justify their responses and state which visualisations they used.

Inspection of the Session Videos

As a basis to compare the information inferred by the facilitators from the visualisations, we inspected the video recordings of each group's session grounding on the theories of collaborative learning of Dillenbourg (1998). Next, we describe the results of these observations.

Group A was highly collaborative from the beginning. They immediately focused on working collaboratively and they built the concept map as a group. They never divided the task. All members of the group discussed every single part and action each group member performed. They worked in parallel for brief periods but never loosing awareness of others' actions. They added key concepts and links, and tried to eliminate redundant concepts by "generalising" them to come up with a clearer map. It is really important to point out that before finishing the trial they realised that the concept map shape should be a circle, given that the water cycle is indeed one. Notably, in the individual maps, none of the participants drew a cyclic map.

Group B worked independently from the beginning and then started collaborating at the second half of the session. At the beginning, they added many concepts and links to the tabletop, explaining and giving brief comments to the others about their actions, which concepts they considered important and asking if they had already added specific concepts that they may use. The concept map at the tabletop, as shown at *left* of Figure 5, is itself a visual aid because it highlights the different propositions added by each participant. This feature proved to be really useful for this group. They divided the task and worked in parallel without collaborating until minute 20. At minute 21 one member of the group expressed: "*your area has more green colour, mine is purple and yours more yellow ... we really came up with three distinct parts. There should be more links between them*". They tried to collaborate after minute 20. The evolution diagram shows this change in the group's behaviour (*right* of Figure 4). Then they worked together to decide which links could best connect the different three main areas of the concept map. However, they ran out of time and the final map was complex and was a poor response to the task.

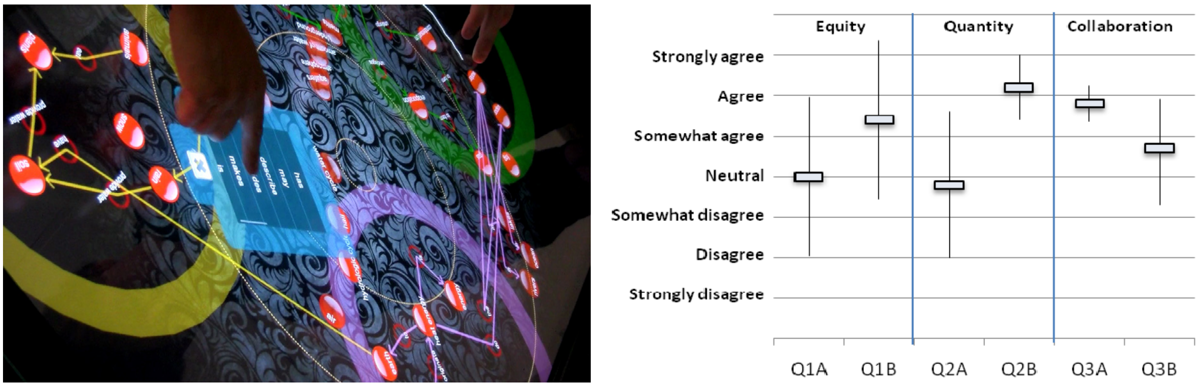


Figure 5. (*left*) The coloured links of the map as a visual indicator of contribution itself. (*right*) Overview of the results of the questions related to: equity of participation (Q1A for group A and Q1B for group B), quantity of participation (Q2A, Q2B) and collaboration (Q3A, Q3B).

Results

All the facilitators were able to easily understand the visualisations and complete the questionnaire without difficulty. Participants were highly engaged in the inspection of the visualisations, and expressed their thoughts verbally, permitting the experimenter to take note of which visualisations influenced their comments and answers about the groups of learners.

The facilitators neither watched the videos nor had access to any summary of what happened during the group sessions. They inspected the visualisations carefully before giving an answer to each posed question for both groups, A and B. To look for an explanation of the acceptance or rejection of each hypothesis, we validated the responses to the questions at two levels. Firstly, we refute a hypothesis if its respective question could not be answered by two or more facilitators. This filter refutes a hypothesis if the facilitators could not find evidence from the visualisations to give an answer. Then, the next step was to ascertain whether their answers matched the observations of video recordings of the group sessions. Table 1 summarises the questionnaire responses. Columns Q1, Q2 and Q3 were used to generate the graph of Figure 5 (*right*). Table 2 summarises which visualisations the facilitators used to give an answer to each question.

(H1) *These visualisations provide useful information about the equity in the roles and participation of the group members*. The focus of this hypothesis is to assess whether the visualisations portray the symmetry of participation of the group members (Dillenbourg, 1998). All facilitators used both the radars of voice and touch together with the contribution charts to answer this question (see Table 2). For group A, the participation was somewhat equitable but one participant slightly dominated most of the verbal participation. This can be observed in Figure 2. In Figure 5 (*right*, Q1A) we notice that in general the responses of the facilitators did not confirm that the group was neither symmetric nor asymmetric. In general, the facilitators judged group B to be symmetric (*right* of Figure 5, Q1B). In fact, the video recordings showed that group B members worked in parallel and did not influence each other. Thus, we accept the usefulness of the radars of physical and audible participation based on the direct observations of the facilitators who remarked that the coloured shaded radars and the concentric circles were useful to quickly detect if the participation of a group was symmetric or not.

Table 1: Summary of the questionnaire responses. Columns: questions asked to the evaluators. Rows: AVG: Average of 7-point Likert scale answers. STD: Standard deviation. * Not all facilitators answered Q4.

Group A	Q1	Q2	Q3	Q4	Q5
	Equity	Quantity	Collaboration	Contribution	Process
AVG	4	3.8	5.8	6	6.25
STD	1.95	1.8	0.45	*	0.5
Lickert	neutral	neutral	agree	agree	agree
Group B					
AVG	5.4	6.2	4.5	6	6.25
STD	1.95	0.8	1.2	*	0.5
Lickert	agree	agree	somewhat agree	agree	agree

(H2) *These visualisations provide useful information about quantity of participation of the group members*. The focus of this hypothesis is to assess if the visualisations depict the quantity and equity of participation of the group members; therefore, it is also related to H1. However, in this case, three of five facilitators used the Contribution Chart in addition to the radars of Participation and the Map Evolution Diagram to answer this question. Group A was reluctant to do physical actions compared with Group B. They focused more on the discussion and negotiation of each proposition to be created and each element to be deleted. By contrast, Group B divided the task and added a large number of propositions. Indeed, the facilitators strongly agreed that in general group B members highly participated in an equitable way. For group A they noticed that the members did not perform many actions and some evaluators expressed that they were talking too much and physically working very slowly (see *right* of Figure 5, Q2A and Q2B). Therefore, we accept the usefulness of the map Contribution Chart in conjunction with the Radars of Participation and the Evolution Diagram because they gave insights to facilitators to size the quantity of participations of the group of learners.

(H3) *These visualisations provide useful information about the group in terms of collaboration.* The focus of this hypothesis is to assess if the visualisations can offer hints to the facilitators to indicate whether the group was collaborative or not. To answer the question related to this hypothesis, the facilitators used all the visualisations including the final product map. All evaluators agreed that group A was collaborative, even though their final product was a small concept map. Moreover, the facilitators considered that in this group the participants interacted with others based on the high levels of talking observed in the verbal radars and the sparse *add link* events shown in the Evolution Diagram (*left* of Figure 4). In the case of Group B three of five facilitators concluded that the group divided the work most of the time given the low levels of talking and the creation of many links in a short time window (*right* of Figure 4). Moreover, the facilitators agreed that the group tried to collaborate in some way before the end of their activity (see *right* of Figure 5, Q3B). We can conclude that the visualisations aid in the perception of various forms of collaboration; however, these results should be confirmed with more case studies.

(H4) *These visualisations provide useful information about the equity of intellectual contribution of the group members*. The purpose of this hypothesis is to assess if the facilitator considers that the content of the concept map reflects intellectual contribution of each member of the group. We refute this hypothesis given the fact that just one of the five facilitators responded to the question related to it (see Table 2). All of them tried to infer intellectual contribution based on the Contribution Chart and the Radar of Voice participation, but afterwards, they concluded that it is difficult to infer the intellectual contribution without knowing the content of the utterances. Indeed, new knowledge is created through the content of the discourse of the group (Stahl, 2006).

(H5) *These visualisations provide useful information about the creation process of the map in terms of the relative contributions of group members and the effectiveness of the group work*. We validated this hypothesis by the direct answers to the corresponding question. All facilitators somewhat agreed with the usefulness of the visualisations because they provide information about the process of map creation in terms of the physical and verbal actions.

Ultimately, the facilitators could answer four of five questions. These give insights into the usefulness of such tools for monitoring the collaborative situation at the tabletop even when the facilitator has the opportunity to observe the actions of the group in situ. Thus, the visualisations aim to complement the qualitative function of facilitators by providing quantitative insights about the group. Even when the results match well with the qualitative observations on equity, quantity, and the most important, the presence of real collaboration, the goal of the visualisations is to aid the qualitative assessment, not to take its place (Stahl, 2006).

Table 2: Visualisations that were used for the facilitators to answer each question.

	Evaluators				
Questions	1	2	3	4	5
1	Radars - Chart	Radars – Chart	Radars - Chart	Radars - Chart	Radar (t) - Chart
2	Radars - Chart	Radars – Chart	Radars	Evolution map	Radars - Chart
3	Final product - Evolution map	Radars – Chart	Radars	Final product	Radars
4	none	Chart	none	none	none
5	Radar (v) - Evolution map	Radars	none	All	Radars

Conclusion

We have presented a set of visualisations that externalise the activity of groups working together at the tabletop to build a group artefact. The Participation Radars provide a mirror of learners' actions both verbal and physical. The Contribution Chart gives an indication of the extent of each learner's contributions to the group artefact. The Evolution Diagram depicts the building process of the artefact relating this to a master artefact and with each participant's contribution. These visualisations trace different aspects of a group, and the combination of the three visualisations with the final group artefact can help the facilitator better understand the nature of the collaboration in order to make improved decisions regarding the guidance of the group of learners.

The ultimate goal of this line of research is placed in the context of fostering co-located collaborative knowledge building tackled through the study of the digital footprints of the learners. Future research will focus on evaluating these visualisations outside the laboratory, in real learning contexts. We also will explore ways to identify the significant patterns of interactions. Given the complexity of group interactions, we believe that it is really important to link the quantitative indicators with other forms of empirical assessment of collaborative accomplishments.

References

Bachour, K., Kaplan, F., & Dillenbourg, P. (2010). An Interactive Table for Supporting Participation Balance in Face-to-Face Collaborative Learning. *IEEE Transactions on Learning Technologies*, 3(3), 203-213

Buisine, S., Besacier, G., Najm, M., AmézianeAoussat, & Vernier, F. (2007). Computer-supported creativity: evaluation of a tabletop mind-map application. In *Proceedings of EPCE 2007*, (pp. 22-31).

Dillenbourg, P. (1998). What do you mean by 'collaborative learning'? Collaborative Learning: Cognitive and Computational Approaches. *Advances in Learning and Instruction Series.* (pp. 1-19): Elsevier Science.

Do-Lenh, S., Kaplan, F., & Dillenbourg, P. (2009). Paper-based concept map: the effects of tabletop on an expressive collaborative learning task. In *Proceedings of BCS-HCI 2009*, (pp. 149-158)

Donath, J. (2002). A semantic approach to visualizing online conversations. *Communications of the ACM*, 45(4), 45-49

Engelmann, T., & Hesse, F. (2010). How digital concept maps about the collaborators' knowledge and information influence computer-supported collaborative problem solving. *IJCSCL, 5(3),* 299-319

Erickson, T., Smith, D., Kellogg, W., Laff, M., Richards, J., & Bradner, E. (1999). Socially translucent systems: social proxies, persistent conversation, and the design of "babble". In *Proceedings of SIGCHI 1999*, (pp. 72-79).

Falcao, T. P., & Price, S. (2009). What have you done! the role of 'interference' in tangible environments for supporting collaborative learning. In *Proceedings of CSCL 2009,* (pp. 325-334)

Harris, A., Rick, J., Bonnett, V., Yuill, N., Fleck, R., Marshall, P., & Rogers, Y. (2009). Around the table: are multiple-touch surfaces better than single-touch for children's collaborative interactions? In *Proceedings of CSCL 2009,* (pp. 335-344)

Janssen, J., Erkens, G., Kanselaar, G., & Jaspers, J. (2007). Visualization of participation: Does it contribute to successful computer-supported collaborative learning*? Computers & Education, 49(4),* 1037-1065.

Jermann, P., Zufferey, G., Schneider, B., Lucci, A., Lepine, S., & Dillenbourg, P. (2009). Physical space and division of labor around a tabletop tangible simulation. In *Proceedings of CSCL 2009,* (pp. 345-349)

Kay, J., Maisonneuve, N., Yacef, K., & Reimann, P. (2006). The Big Five and Visualisations of a Team Work Activity. In *Proceedings ITS 2006,* (pp. 197-206) .

Martinez, R., Kay, J., & Yacef, K. (2010). Collaborative concept mapping at the tabletop. In *Proceedings of Interactive Tabletops and Surfaces 2010,* (pp. 207-210).

Morgan, M., & Butler, M. (2009). Considering multi-touch display technology for collaboration in the classroom. In *Proceedings EDMEDIA 2009*, (pp. 674–683)

Novak, J., & Cañas, A. (2008). The Theory Underlying Concept Maps and How to Construct and Use Them. *In TRIC 2006-01* (Ed.): Florida Institute for Human and Machine Cognition.

Oppl, S., & Stary, C. (2009). Tabletop concept mapping. In *Proceedings of* TEI 2009. (pp. 275-282)

Paavola, S., & Hakkarainen, K. (2005). The Knowledge Creation Metaphor – An Emergent Epistemological Approach to Learning. *Science & Education, 14(6),* 535-557.

Race, P. (2001). A briefing on self, peer & group assessment. *Assessment Series N. 9.* Learning and Teaching Support Network.

Stahl, G. (2006). Group Cognition: Computer Support for Building Collaborative Knowledge: MIT Press

Sugimoto, M., Hosoi, K., & Hashizume, H. (2004). Caretta: a system for supporting face-to-face collaboration by integrating personal and shared spaces. In *Proceedings* SIGCHI. (pp. 41-48)

Sundararajan, B. (2010). Emergence of the Most Knowledgeable Other (MKO): Social Network Analysis of Chat and Bulletin Board Conversations in a CSCL System. *Journal of e-Learning* 8(2), 191-208.

Suthers, D., Vatrapu, R., Medina, R., Joseph, S. & Dwyer, N. (2008). "Beyond threaded discussion: Representational guidance in asynchronous collaborative learning environments." *Computers and Education*. 50(4), 1103-1127.

Taricani, E., & Clariana, R. (2006). A Technique for Automatically Scoring Open-Ended Concept Maps. *Educational Technology Research and Development, 54(1)*, 65-82.

Tergan, S.-O. (2005). Digital Concept Maps for Managing Knowledge and Information. *Knowledge and Information Visualization*. pp. 185-204

Upton, K., & Kay, J. (2009). Narcissus: Group and Individual Models to Support Small Group Work, In *Proceedings of UMAP 2009* (pp. 54-65).

Acknowledgments

This research was funded by the Smart Services Cooperative Research Centre, Australia. We thank Professor Peter Reimann for his suggestions and the experiment participants for their time.

Questioning and the Quality of Knowledge Ideas in a CSCL Context: A Study on Two Age-groups of Students

Ming Lai, Nancy Law, The University of Hong Kong, Pokfulam Road, Hong Kong
Email: minglai@hkucc.hku.hk, nlaw@hku.hk

Abstract: This paper aims to study the relationship between the type of questions posted and the quality of knowledge ideas expressed in a CSCL context. Previous studies suggest that this relationship may be different for different age-groups. This study was conducted with two classes of students—at grades 6 and 10 respectively. The discourse was analyzed at both thread- and individual-student- levels. It was found that for the sixth-grade students, those who asked better questions also expressed better knowledge ideas. For the tenth-grade students, such a relationship at the individual-student-level was not found. Nonetheless, at the thread-level, there is a positive correlation between the presence of better questions and more developed knowledge ideas in the tenth-grade students' online discourse, but not in the sixth-grade students'. The results suggest further discourse characteristics in addition to questioning to be related to the quality of knowledge ideas expressed.

Introduction

The relationship between questioning and inquiry has been articulated since the time of ancient Greeks. Socrates illustrates how to use questions to seek for knowledge. Aristotle suggests that the kinds of questions we ask are directly related to the knowledge we can get and he is the first person to develop a classification for different kinds of questions (Dillon, 1984). In the context of computer-supported collaborative learning (CSCL), students can conduct collaborative inquiry in a technology-mediated environment. It is important for them to generate questions and to build their inquiry on these questions (Scardamalia & Bereiter, 1991). The pedagogical approach of knowledge building emphasizes students' collective responsibility for the advancement of the community's knowledge (Scardamalia, 2002). Its implementation in schools is usually integrated with an asynchronous online platform, the Knowledge Forum® (KF), which is specifically designed to create a knowledge building environment for students to make their ideas explicit so that everyone can contribute to the continual improvement of ideas (Scardamalia, 2002). Student's self-generated questions are considered especially important in knowledge building as it emphasizes the epistemic agency of students (Scardamalia, 2002; Zhang, Scardamalia, Lamon, Messina, & Reeve, 2007). On the other hand, students can articulate their "knowledge ideas" (Hakkarainen, 2003) in the process of inquiry, which can be intuitive ideas, scientific theories, or the information they gathered. Based on these knowledge ideas, students can generate further questions so that their inquiry can be carried forward (Hakarainen, 2003). Hence, during the course of an inquiry, there will be a number of questions as well as knowledge ideas generated by students. Although there are case studies reporting that by formulating better questions, students can articulate better knowledge ideas, less research is conducted to examine the relationship between questioning and the quality of knowledge ideas quantitatively. In two previous studies in which a student was taken as the unit of analysis, some contradictory results on the above relationship were found (Hakkarainen, Lipponen, & Jarvela, 2002; Lee, Chan, & van Aalst, 2006). Nonetheless, to capture a complete inquiry, a discussion thread seems to be more appropriate as the unit of analysis for examining the relationship between questioning and knowledge ideas. This paper aims to study the relationship at both thread- and individual-student- levels. As suggested by previous studies, the relationship might be different for different age-groups (Hakkarainen et al., 2002; Lee et al., 2006), a total of two age-groups of students are examined in this study.

Literature Review

In traditional classrooms, questions are usually asked by teachers. The structure of initiation-response-evaluation (IRE, Mehan, 1979) depicts the patterns often found in classroom discourse, in which the teacher initiates a question; the students respond; and then the teacher evaluates the responses. The number of questions students asked in classroom settings is low (Graesser & Person, 1994). On the other hand, there are studies reporting that students' questioning is beneficial to their text comprehension and learning (e.g., King, 1990; Palincsar & Brown, 1984); it is also found that even elementary students are capable of generating questions of high quality (Hakkarainen, 2003; Scardamalia & Bereiter, 1991). Compared to an ordinary classroom, collaborative inquiry pedagogical designs allow students to generate questions for further inquiry more easily. In CSCL studies, the analysis of discourse related to questions is often found (e.g., Hakkarainen, 2003; Hmelo-Silver, 2003; Zhang et al., 2007).

To analyze the questions generated by students, it is necessary to classify different types of questions that can be found in educational settings. In the past years, different classification frameworks have been developed (e.g., Dillon, 1984; Grasser, 1994; King & Rosenshine, 1993). For example, in analyzing the questions students asked in interactions related to a task of text comprehension, King and Rosenshine (1993) differentiated the following question types in order of increasing quality, 1) fact questions; 2) definition questions; and 3) integration questions targeting the integration of ideas or information. In their study, King and Rosenshine (1993) also investigated the effect of training students to ask questions with the help of question stems. They found that compared to the control group, students in the training condition asked more high-quality questions and the knowledge they constructed was also of higher quality (see also King, 1994).

Students' questions are also classified in the literature on knowledge building, and the classification is based on their epistemological nature—whether they target factual or explanatory knowledge. Questions seeking explanations are more productive from the perspective of knowledge building as they are related to a deeper level of understanding compared to fact-seeking ones (Hakkarainen, 2003; Lee et al., 2006; Zhang et al., 2007). In addition to the questions asked, students can articulate knowledge ideas in the process of inquiry, which can be responses to questions, intuitive theories, and scientific information (Hakkarainen, 2003). In the literature on knowledge building, the quality of knowledge ideas is examined according to their "levels of explanation" (e.g., Hakkaranen, 2003; Lee et al., 2006; Zhang et al., 2007). The rationale is again based on the epistemological nature of knowledge, in which explanation is more illustrative of deep understanding compared to factual knowledge. A total of four levels of explanation are classified in Zhang et al.'s (2007) study: 1) unelaborated facts; 2) elaborated facts; 3) unelaborated explanations; and 4) elaborated explanations, indicating an increase in quality. Fact refers to the description of terms, phenomena, or experiences, while explanation refers to reasons, relationships, or mechanisms. In addition to this differentiation of fact and explanation, Zhang et al. (2007) takes into consideration how elaborated the knowledge ideas are.

Both students' questions and their knowledge ideas are important elements in the process of inquiry, as questions may lead to the articulation of knowledge ideas, and the articulation of knowledge ideas may lead to further research questions. Through continuous iterations of these cycles, students' collective knowledge can be advanced. Case studies have been reported to describe how elementary students can engage in these iterated cycles to advance their inquiry (Hakkarainen, 2003; Zhang et al., 2007). Although case studies are helpful to illustrate how students may deepen their inquiry by questioning and putting forward knowledge ideas, a correlation analysis may be a more cogent way of investigating the relationship between levels of questions and levels of explanations. Hakkarainen et al. (2002) analyzed this relationship among three classrooms of fourth-grade and fifth-grade students, finding a positive correlation. However, the unit of analysis they employed was a student, thus more accurately, the correlation implied that a student generating a higher proportion of explanation-seeking questions was also the one who has a higher level of explanation expressed. In another study with a sample of ninth-grade students, Lee et al. (2006) found that the correlation between levels of question and levels of explanation was not significant. Nonetheless, both of these studies employed a student as the unit of analysis. Although a thread seems to be more appropriate to capture a complete inquiry as it contains a series of related notes, in the literature on knowledge building, no research has been attempted to examine the relationship between questioning and levels of explanation quantitatively with a thread as the analytic unit. In this study, both a thread and a student are taken as the units of analysis respectively to examine the relationship between the types of questions and the quality of knowledge ideas.

Method

Research Context

The data analyzed in this study were part of the "Learning Community Project" (LCP), which was launched to promote knowledge building and support its implementation in secondary and primary schools in Hong Kong. As previous studies suggest, the relationship between levels of questions and levels of explanation may be different for different age groups (Hakkarainen et al., 2002; Lee et al., 2006), a total of two databases, one primary and one secondary, were analyzed to see if different patterns were found. Previous studies suggested that a long period of inquiry is needed for productive knowledge building discourse to emerge (e.g., Zhang et al., 2007), these two databases had the longest period of inquiry among all databases at the time the analysis was conducted. Students of both grades had a period of about six weeks to conduct their inquiry.

The first database contains the discourse of a class of 41 tenth-grade students in a secondary school; they formed six groups inquiring on topics of Water Quality, Plastics, and Ideal Vehicle. The students and their teacher were both new to knowledge building. The second database composed of the discourse of a class of 44 sixth-grade students in a primary school, who formed seven groups to inquire on topics of Global Warming, Energy Crisis, and Species Extinction. Two teachers were involved in the facilitation of knowledge building in the second database. One of them and about one fourth of the students had the experience of participating in knowledge building activities in the previous year. The other teacher and the remaining students were new to this pedagogical approach.

Knowledge Forum®

Throughout the process of inquiry, students had to engage in discussions on Knowledge Forum® (KF), which is an asynchronous online platform specifically designed to facilitate knowledge building (Scardamalia, 2003). Students can contribute their questions and ideas in the form of notes, and other students could respond and further improve the ideas by writing build-on notes. In writing a note, students may use “scaffolds”, which are meta-cognitive prompts in the form of word cues such as “New information”, “New idea”, and “My theory”, so that they can better identify the nature of their note content (e.g., “New information” or “New idea”), or identify a knowledge gap (e.g. “I need to understand”). Notes and their build-on notes were linked physically in the form of build-on threads. Thus within a thread, there were a series of related notes.

Levels of Questions and Levels of Explanation

The online discourse collected in this study is analysed to identify, for each note, the quality of questions asked and the level of explanation of knowledge ideas expressed (where relevant). “Explanation-seeking questions” refer to those targeting at explanations, while “fact-seeking questions” are those targeting at factual information (Hakkarainen, 2003). In addition to these two categories, there were questions targeting at simple clarifications, hence a third category, “simple clarification”, was included. Such a differentiation was adopted in a recent study conduct by van Aalst (2009). Presented in table 1 are examples of these three levels of questions taken from students’ notes collected in this study.

Table 1: Examples of the three levels of questions generated by students.

Levels of Questions	Examples
Explanation-seeking	“How UV (ultra violet) works to improve water quality?”
Fact-seeking	“Which country is using solar energy to move the car?”
Simple clarification	“Can you put up some picture, to let us know what red panda looks like?”

In parallel to the classification of levels of questions, students’ non-question-asking notes were analyzed according to levels of explanation of their knowledge ideas. The analysis involves two steps: 1) determining whether a knowledge idea is articulated in a note; and 2) if a knowledge idea is articulated, what its level of explanation is. To be qualified as a “knowledge idea”, some knowledge contents have to be expressed in a note. The following are examples of notes with no knowledge idea articulated: “*I agree*”, “*We can have a further discussion on this topic*”. Notes containing simply copy-and-paste contents, information of websites, or pictures copied from elsewhere were also considered as having no knowledge ideas expressed.

For notes with knowledge ideas expressed, the coding scheme developed by Zhang et al. (2007) was employed to identify their levels of explanation. Presented in table 2 are examples of notes coded at these four levels in this study.

Table 2: Examples of the four categories of levels of explanation.

Categories	Examples
Elaborated explanation	“UV (Ultra Violet) is a light wave which has more energy than the visible light. Its wave length is shorter so that every time it contains more energy. This energy can change the nature of the bacteria so the bacteria die.”
Unelaborated explanation	“UV (Ultra Violet) is a light. It is a kind of waves and it is not a matter. It does not remain in the water.”
Elaborated fact	“Fossil fuel is composed of three kinds of elements: gas, oil, or coal. Fossil fuel energy is a nonrenewable type of energy, that means, it would disappear if we use it all or if we waste it, we wouldn’t get it again. Fossil fuels are the remains of ancient plant and animal life found in earth, rock, and clay. Fossil fuels are mined by people for use as an energy source.”
Unelaborated fact	“In fossil fuel, there are coal and oil.”

To conduct the quantitative analysis, the coding of the level of explanation was converted to a numeric score according to the procedure in Zhang et al.'s (2007) study: “elaborated explanation” was assigned with a score of “4”, “unelaborated explanation” as “3”, “elaborated fact” as “2”, and “unelaborated fact” as “1”, indicating a decrease in quality. Notes with no knowledge ideas expressed were not assigned with any score. The “average level of explanation” of a build-on thread was calculated by averaging the levels of explanation of the notes within the thread. On the other hand, the “average level of explanation” of a student was calculated by averaging the levels of explanation of the notes contributed by the student.

Similarly, for questions generated by students, an explanation-seeking question was assigned a score of “3”, a fact-seeking question “2”, and a simple clarification question “1”, indicating a decrease in quality. The “average level of questions” of a thread was calculated by averaging the scores of questions generated in a thread. Similarly, the “average level of questions” of a student was calculated by averaging the scores of questions contributed by the student. While the quality of questioning was captured by the “average level of questions”, a quantitative measure was also included in this study. The “questioning density” of a thread was calculated by dividing the number of notes with questions asked divided by the total number of notes within the thread—it is hence the proportion of notes in a thread containing questions. Similarly, the "questioning density" of a student was calculated by dividing the number of notes with questions by the total number of notes contributed by the student; it was a quantity measure of how many questions were asked by a student with respect to the total number of notes he/she created.

Results

Descriptive Analysis: Levels of Questions and Levels of Explanation

A total of 620 and 630 discussion notes were generated by the tenth-grade and sixth-grade students respectively, which resulted in a total of 76 and 69 build-on threads respectively. Among the notes created, a total of 231 and 132 questions were identified in the discussion of the tenth-grade and sixth-grade students respectively, suggesting that the tenth-graders tended to ask more questions compared to the sixth-graders. All these questions were analyzed on their levels of questioning. Presented in table 3 are the numbers and percentages of questions classified at different levels. Nearly half of the questions of both grades of students were classified as explanation-seeking. Only a low percentage of their questions were simple clarifications (Tenth-grade: 15.6%; Sixth-grade: 20.5%). The tenth-graders had about 40% of their questions classified as fact-seeking, while that for the sixth-graders was about 30%. Hence although the two grades of students differed in the numbers of questions asked, the proportions of levels of questions were quite similar.

Table 3: Numbers and percentages of questions in different levels of questions.

Grade		Levels of Questions			
		Simple clarification	Fact-seeking	Explanation-seeking	Total
Tenth	Number	36	92	103	231
	%	15.6	39.8	44.6	100
Sixth	Number	27	41	64	132
	%	20.5	31.1	48.5	100

On the other hand, a total of 430 and 425 notes of the tenth-grade and sixth-grade students respectively were found to contain knowledge ideas, and these were analyzed for their levels of explanation. Table 4 presents the distribution of levels of explanation of the notes with knowledge ideas. It can be seen that for both grades, slightly more than half of their notes with knowledge ideas were classified as unelaborated facts (Fact-Unelab.). 17% and 16.9% of the notes with knowledge ideas of the tenth-grade and sixth-grade students respectively were elaborated facts (Fact-Elab.). The tenth-grade students had 27.9% (20+7.9) of their notes with knowledge ideas classified as explanatory, which could either be unelaborated explanations (Expl.-Unelab.) or elaborated explanations (Expl.-Elab,), comprising a total of 120 notes. For the sixth-grade students, 28.7% (19.1+9.6) of their notes with knowledge ideas were classified as explanatory, which comprised a total of 122 notes.

Table 4: Distribution of notes with knowledge ideas in different levels of explanation.

Grade		Levels of Explanation				
		Factual		Explanatory		
		Fact-Unelab.	Fact-Elab.	Expl.-Unelab.	Expl.-Elab.	Total
Tenth	Number	237	73	86	31	430
	%	55.1	17	20	7.9	100
Sixth	Number	231	72	81	41	425
	%	54.4	16.9	19.1	9.6	100

Correlation Analysis

Correlation with a Thread as the Unit of Analysis

Previous literature on knowledge building found that short threads are less illustrative of a deeper level of understanding (Zhang et al., 2007). In this study, there were threads with only a small number of notes; they are deemed to be too short to illustrate how questioning and levels of explanation are related within a thread. Hence only threads with six notes or more are included in the analysis in this section. A total of 45 (59.2% out of 76) and 38 (55.1% out of 69) build-on threads of tenth-grade and sixth-grade students respectively were with six notes or more, and they were included in the correlation analysis. The correlation results are summarized as in table 5. For the tenth-grade students, the average level of explanation of a thread was significantly correlated to the average level of question ($r=.448$, $p<.01$) but not questioning density of a thread, suggesting that the quality rather than quantity of questions was related to the level of explanation of a thread for the tenth-graders. For the sixth-grade students, neither the average level of questions nor questioning density was significantly related to the average level of explanation of a thread.

Table 5: Relationship between levels of explanation and questioning with a thread as the analytic unit.

Grade		Average Levels of Questions	Questioning Density
Tenth (n=45)	Average Level of Explanation	.448**	.026
	Average Level of Questions		.161
Sixth (n=38)	Average Level of Explanation	.295	-.245
	Average Level of Questions		.284

*: p<.05; **: p<.01

Correlation with a Student as the Unit of Analysis

If a student is taken as the unit of analysis rather than a thread, a contrasting correlation pattern was found. As summarized in table 6, no significant correlations between the average level of explanation of a student and the average level of questions or questioning density were found for the tenth-graders. On the contrary, the average level of explanation of a student was positively correlated to the average level of questions for the sixth-graders. Such findings were in fact consistent with previous research that levels of questions were found to be related to levels of explanation for primary grade but not secondary grade students if a student was taken as the unit of analysis (Hakkarainen et al., 2002; Lee et al., 2006).

Table 6: Relationship between levels of explanation and questioning with a student as the analytic unit.

Grade		Average Levels of Questions	Questioning Density
Tenth (n=41)	Average Level of Explanation	-.174	-.167
	Average Level of Questions		.2
Sixth (n=44)	Average Level of Explanation	.425*	.054
	Average Level of Questions		.074

*: p<.05

Case Studies

In order to make sense of the above findings in the context of this study, two case studies are presented in this section. The first case is a sixth-grade student who is among the highest in both the level of questions and the level of explanation expressed. The second case is a thread of the tenth-grade students, which illustrates how questioning might deepen the level of explanation in a thread; it also illustrates that students asking better question might not be the ones who contributed notes with higher levels of explanation.

Sixth-grade Students

Tommy (fake name) was a sixth-grade student. He has written a total of 21 notes, distributed among eleven threads. Of the notes he has written, nine contained explanatory knowledge ideas (unelaborated or elaborated explanation) and a total of five questions were asked, of which four were explanation-seeking ones. His explanation-seeking questions distributed among three threads, yet they did not seem to be helpful to advance the inquiry in those threads. For example, in the thread of "Ground heat energy", in response to the articulation that not all places are suitable for this form of energy, Tommy asked the question, "*But how come New Zealand can do it?*" The question was not addressed by other students at all. In the thread of "Solar energy", although Tommy did not ask any questions, he contributed two notes with unelaborated and elaborated explanations respectively, as illustrated in the contents of notes taken from this thread presented in table 7. In this thread, students were arguing on whether solar energy is useful. In the first note written by Tommy (Note#A4), it was said that solar energy is not that useful because there may be rainy days. Then a student opposed to this point by articulating that the weather of Hong Kong, the place they live in, is usually hot. In the remaining notes, Tommy tried to clarify whether it is the hot weather or sunlight that generates solar energy. Explanatory-seeking questions were also found in this thread (Note#A13 & A14), but they received no responses.

Table 7: Contents of notes taken from the thread of "Solar energy" of the sixth-grade students.

Note#	Student	Note Content	Coding
A1	G6_1	Solar energy is good because1.All countries have sun 2.Solar energy does not need too much space.	Fact-Unelab.
A2	G6_2	I agree your own opinion, cause it won't need a lot of space and everywhere got sunshine.	Fact-Unelab.

A3	G6_3	I agree too because it use less energy.	Fact-Unelab.
A4	**Tommy**	**I don't think solar energy is good because when there are some continuous rainy days in a week, the day time might not be long enough to produce energy, so if we rely on the solar energy, it might cause no electricity for us to use in summer.**	Fact-Elab.
A5	G6_4	I think it is not a good point. Because Hong Kong is always very hot.	Fact-Unelab.
A6	**Tommy**	**There might be hot days, but I'm not sure that there is sunlight. Solar energy needs sunlight.**	Expl.-Unelab.
A7	G6_2	I'm sure that everyday might be sunlight, unless there's bad weather, like typhoon, raining heavily and also storm......	Fact-Elab.
A8	G6_3	I think the solar energy machine can get the energy and use for more than one day.	Expl.-Unelab.
A9	G6_4	Hong Kong is a very hot place.	Fact-Unelab.
A10	G6_3	Hong Kong is not EVERYDAY hot. Sometimes, HK can be cold	Fact-Unelab.
A11	**Tommy**	**Hot doesn't mean that there is enough sunlight, you need sunlight to heat up the solar panels. Just like today, although it is hot, there is no sunlight. While in Arctic, there are some extreme situations, like 10 days have sunlight, and 10 days NO sunlight.**	Expl.-Elab.
A12	G6_4	Solar energy can help to solve problems for people around the world. In some places, people do not use electric or gas cookers.	Fact-Unelab.
A13	G6_1	How can the sun help them cook?	*Quest.-Expl.*
A14	G6_3	What is the way of fixing the solar panels if the panels are out of order? It is a very big problem. We will have no energy to use!!!	*Quest.-Expl.*

Tenth-grade Students

In the thread of the sixth-grade students presented in the last section, questioning seemed to play a limited role in the advancement of knowledge. For the tenth-grade students, as suggested by the correlation results, there seemed to be a positive relationship between levels of questions and levels of explanation of a thread. Table 8 presents the note contents written in the thread of "Substance of fuels". The thread began with a fact-seeking question of "What is the chemical substance in fossil fuel?" It received an unelaborated factual response that fossil fuel contains coal and oil. Then the students went on with the inquiry on other substances which can be sources of fuel such as LPG (Liquefied Petroleum Gas) and water. The proposal led to the request for explaining how water can be a kind of fuel, which resulted in the inquiry on the mechanism of fuel cells, in which liquid hydrogen is used for generating electricity with water as the byproduct (Note#B11). In the last few notes, the students inquired on whether the output of a fuel cell is strong enough.

Throughout this thread, a series of questions were generated to drive the inquiry forward, and the discourse changed from factual (Note#B1 to B3) to explanatory (Note#B4 to B11). A total of eight questions were found in this thread. Although the students did not start their inquiry with an explanation-seeking question, two questions of this type were generated in the process of inquiry, leading to the explanatory exploration of how water can be a source of fuel and the mechanism of fuel cell. It can also be observed from this thread that the students who contributed notes with unelaborated or elaborated explanations (G10_3. G10_6, G10_2) are not the same as those who asked explanation-seeking questions (G10_4, G10_1) or other types of questions (G10_5, G10_7). Such a finding is in line with the previous result that there was no significant correlation between levels of questions and levels of explanation at an individual level for the tenth-grade students.

Table 8: Contents of notes taken from the thread of "Substance of fuels" of the tenth-grade students.

Note#	Student	Note Content	Coding
B1	G10_1	What is the chemical substance in fossil fuel?	*Quest.-Fact*
B2	G10_2	In fossil fuel, there are coal and oil.	Fact-Unelab.
B3	G10_3	To protect the environment, I think using LPG is better............. Also, some Japanese scientist tries to use water to be fuel.........	Fact-Unelab.
B4	G10_4	Water..? Is it really possible?	*Quest.-Expl.*
B5	G10_3	Water is formed by hydrogen and oxygen. They could be a kind of good fuel.	Expl.-Unelab.
B6	G10_5	How????	*Quest.-Sim.*
B7	G10_6	Decompose water, form hydrogen and oxygen, then burn both of them.	Expl.-Unelab.
B8	G10_1	Burning hydrogen and oxygen can supply such great energy for the car to move?	*Quest.-Expl.*
B9	G10_4	I think it is the problem of the amount you use only...	Fact-Unelab.

B10	G10_6	Of course is enough! Hydrogen burn in air and oxygen support combustion. So they can provide that much energy.	Expl.-Unelab.
B11	G10_2	Fuel cells have been developed which convert hydrogen directly into electricity. This is attractive since the only byproduct is water. There are still significant problems since carbon dioxide is typically produced by use of electricity, which is mostly produced from fossil fuels.	Expl.-Elab.
B12	G10_1	Sorry, what is fuel cell? Is it a kind of fuel that very common to use in car?	*Quest.-Fact*
B13	G10_7	Fuel cell: An electrochemical cell in which the energy of a reaction between a fuel, such as liquid hydrogen, and an oxidant, such as liquid oxygen…. source: http://dictionary.reference.com/search?q=fuel%20cell	Copy & Paste
B14	G10_2	Here is the picture of the fuel cell running process: http://www.newmango.com/info_fuel.html From the picture, we know that the Natural gas + air -> DC POWER and Water+ Heat.	Fact.-Unelab.
B15	G10_7	How about the output of the cell? Will it be strong?	*Quest.-Fact*
B16	G10_2	I am sorry I do not understand what means strong?	*Quest._Sim.*
B17	G10_7	I mean will the current be strong enough?	*Quest.-Fact*
B18	G10_2	Yes, the fuel cell can provide enough energy to the car.	Fact.-Unelab.

Discussion

This paper examines the relationship between levels of questions and levels of explanation at both a thread level and an individual level. At a thread level, a positive correlation was found for the tenth-grade but not the sixth-grade students. An opposite pattern was found at an individual level, in which a positive correlation was found for the sixth-grade but not the tenth-grade students. The latter finding is in fact compatible to previous studies that a positive correlation at an individual level was found for primary grade but not secondary grade students (Hakkarainen et al., 2002; Lee et al., 2006).

If the results of the sixth-grade and tenth-grade students are considered together, there appears to be a developmental difference between younger and older students. For the younger students (sixth-grade), those who ask good questions are the ones with high-level explanations expressed. For the older ones (tenth-grade), the ability to ask good questions appears to have developed independently of the ability to construct high-level explanations. Nonetheless, in a build-on thread of the tenth-graders, there were good questions asked by some students and high-level explanations constructed by other students, resulting in a positive correlation between levels of questions and levels of explanation of a thread. On the contrary, in the discussion of the sixth-grade students, good questions were found in some threads and yet the students addressing them were not able to construct high-level explanations; while in other threads, although good questions were not found, there were notes with high-level explanations.

In the context of CSCL or collaborative learning more generally, students usually engage in collaborative inquiry, and it is important for good questions to be generated and high-level knowledge ideas expressed. More studies are needed to explore whether such a developmental difference really exists, whether it is related to students' ability or their willingness to ask questions and express ideas, whether such a difference is hindering or facilitative to collaborative learning, and whether it can be weakened or strengthened through suitable pedagogical designs.

In previous studies on knowledge building in which the participants were mainly primary grade students (e.g., Hakkarainen, 2003; Hakkarainen et al., 2002; Zhang et al., 2007), it was reported that even elementary students were capable of asking questions of good quality and to build their inquiry on these questions. While the sixth-grade students in this study seemed to be less able to advance their inquiry through questioning. The case study also suggested that even some good questions were asked by the sixth-grade students, they were not addressed by the others. One possible reason is that the teachers and students were relatively new to the pedagogical approach of knowledge building. In previous studies, it was mentioned that the teacher played an important role in the facilitation of knowledge building (Hakkarainen, 2003; Zhang et al., 2007). It seemed that more facilitation from the teachers is needed in helping younger students to formulate better questions, address the questions that were not responded to, and further their inquiry through questioning. However, as suggested by the thread of "Solar energy" presented in the case study, although questioning seemed to play a limited role, some discourse progresses were observed with respect to the oppositions and counter-oppositions put forward by the students, arguing on whether solar energy is useful and whether it can be applied in the context of Hong Kong.

In the literature on knowledge building, it is reported that knowledge building is different from argumentation because the latter focuses on persuasion and the winning of arguments rather than shared understanding and progress (Scardamalia & Bereiter, 2006). However, in more recent studies, argumentation is mainly regarded as a form of discourse found in social contexts when people expound and explore diverse or contradictory ideas (Andriessen, 2006; Leitao, 2000). Rather than focusing on who wins an argument, analytic frameworks are developed to examine how students interact with one another in the presence of opposite ideas (Erduran, Simon, & Osborne, 2004; Leitao, 2000; Pontecorvo & Girardet, 1993). These frameworks might provide valuable insights on how progress can be made through discourse. Although the case study of the tenth-grade students suggested that they were more capable to construct explanations to questions raised by other students, and in fact a significant correlation was found between levels of questions and levels of explanation of a thread for the tenth-grade students, the percentage of variance that can be explained for is about 20% (square of .448), further studies are needed to examine other discourse characteristics such as argumentative interactions that are related to the deepening of explanation within a thread.

References

Andriessen, J. (2006). Arguing to learn. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 443-460). NY: Cambridge University Press.

Bereiter, C. (2002). *Education and mind in the knowledge age*. Mahwah, NJ: Lawrence Erlbaum Associates.

Dillon, J. T. (1984). The classification of research questions. *Review of Educational Research*, 54(3), 327-361.

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the use of Toulmin's Argument Pattern in studying science discourse. *Science Education*, 88(6), 915-933.

Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal*, 31, 104-137.

Hakkarainen, K. (2003). Progressive inquiry in a computer-supported biology class. *Journal of Research in Science Teaching*, 40(10), 1072-1088.

Hakkarainen, K., Lipponen, L., & Jarvela, S. (2002). Epistemology of inquiry and computer-supported collaborative learning. In T. Koschmann, R. Hall, & N. Miyake (Eds.), *CSCL II: Carrying forward the conversation* (pp. 129-156). Mahwah, NJ: Lawrence Erlbaum Associates.

Hakkarainen, K., & Sintonen, M. (2002). The interrogative model of inquiry and computer-supported collaborative learning. *Science & Education*, 11, 25-43.

Hmelo-Silver, C. E. (2003). Analyzing collaborative knowledge construction: Multiple methods for integrated understanding. *Computers & Education*, 41, 397-420.

King, A. (1990). Enhancing peer interaction and learning in the classroom through reciprocal questioning. *American Educational Research Journal*, 27, 664-687.

King, A. (1994). Guiding knowledge construction in the classroom: Effects of teaching children how to question and how to explain. *American Educational Research Journal*, 31, 338-368.

King, A., & Rosenshine, B. (1993). Effects of guided cooperative questioning on children knowledge construction. *Journal of Experimental Education*, 61, 127-148.

Lee, E. Y. C., Chan, C. K. K., & van Aalst, J. (2006). Student assessment of collaborative learning in a CSCL environment. *International Journal of Computer-Supported Collaborative Learning*, 1(1), 57-87.

Leitao, S. (2000). The potential of argument in knowledge building. *Human Development*, 43, 332-360.

Mehan, H. (1979). *Learning lessons: Social organization in the classroom*. Cambridge, MA: Harvard University Press.

Pontecorvo, C., & Girardet, H. (1993). Arguing and reasoning in understanding historical topics. *Cognition & Instruction*, 11, 365-395.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal education in a knowledge society* (pp. 67-98). Chicago: Open Court.

Scardamalia, M. (2003). Knowledge building environments: Extending the limits of the possible in education and knowledge work. In A. DiStefano, K. E. Rudestam, & R. Silverman (Eds.), *Encyclopedia of distributed learning* (pp. 269-272). Thousand Oaks, CA: Sage.

Scardamalia, M., & Bereiter, C. (1991). Higher levels of agency for children in knowledge building: A challenge for the design of new knowledge media. *The Journal of the Learning Sciences*, 1(1), 37-68.

Scardamalia, M., & Bereiter, C. (2006). Knowledge building: theory, pedagogy, and technology. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 97-118). NY: Cambridge University Press.

Stahl, G., Koschmann, T., Suthers, D. (2006). Computer-supported collaborative learning: A historical perspective. In R. K. Sawyer (ed.), *The Cambridge handbook of the learning sciences* (pp.409-426). NY: Cambridge University Press.

van Aalst, J. (2009). Distinguishing knowledge-sharing, knowledge-construction, and knowledge-creation discourses. *International Journal of Computer-Supported Collaborative Learning*, 4, 259-287.

Zhang, J., Scardamalia, M., Lamon, M., Messina, R., & Reeve, R. (2007). Socio-cognitive dynamics of knowledge building in the work of 9- and 10-year-olds. *Educational Technology Research and Development*, 55, 117-145.

Location-based Collaborative Learning at a Geography Trail: Examining the Relationship among Task Design, Facilitation and Discourse Types

Esther Tan, Hyo-Jeong So, Learning Sciences Lab, National Institute of Education, Nanyang Technological University, 1 Nanyang Walk, Singapore
Email: esther.tan@nie.edu.sg, hyojeong.so@nie.edu.sg

Abstract: In this paper, we examine the characteristics of a discourse that show evidences of collective knowledge construction and investigate the impact of task design and facilitation on in situ small group collaborative learning. To examine discourse types, all audio-recorded verbal data of the three groups of secondary students is transcribed, coded and analysed with respect to two key dimensions in the knowledge construction process, namely, *the epistemic* and *the social*. Tasks were categorized largely into *performative* and *knowledge-generative*. Analysis showed that different epistemic activities and the nature of facilitation have a definitive bearing on group discursive moves and more importantly, the presence of a real world context could generate intense knowledge co-construction even for mundane performative tasks. In conclusion, we propose that a three-prong approach (FAT) – Facilitation, Activities in-situ, and Technology – should be considered to support meaningful collaborative learning practices in mobile learning.

Introduction

Rapid advancement in information and communications technology has revolutionised the teaching and learning landscape; creating new and exciting possibilities for learning beyond the four walls of the classroom, and thereby, inevitably changing the role of both the learner and the teacher. Seamless learning across context and space is now made possible with the affordances of mobile computing and web applications. Interaction with the real world environment has also given learning, new shades of meaning and intent. In the field of computer-supported collaborative learning (CSCL), whilst there is extensive research and literature on leveraging emerging mobile devices and Web 2.0 technologies to enhance the learning experience in the real-world setting and/ or learning on the move (e.g., Squire & Klopfer, 2007), there remains little empirical research on other equally significant configurations of mobile learning: that is the *design configuration* of mobile learning environments and the execution path to bring about the desired learning outcomes. Another common pitfall is the focus on the unending possibilities of mobile devices and innovative software applications over the rich affordances of the physical environment in the context of enhancing mobile learning.

To examine the design configurations and the affordances of real-world contexts for promoting meaningful collaborative learning experiences, this paper articulates the in situ learning experience of three groups of Secondary One students (ages 12-13) on a geography mobile learning trail. Specifically, we are interested in investigating the intricate relationship among task-design, facilitation and the discourse types in situated small group collaborative learning. This study analyses the content, process and pattern of students' interactional discourse in the knowledge construction process. We believe that a focus on process rather than outcome is pertinent to understanding the design configurations necessary for effective in situ collaborative learning.

Theoretical Frameworks

The anatomy of location-based collaborative learning primarily composes of two main constructs: *context-oriented* and *a collective enterprise*. More importantly, these two constructs should not be conceived as separate entities operating in their own paradigms; rather they are closely interwoven in the process of knowledge construction to bring about the desired learning outcomes. Educational researchers argue that students learn best when they are able to learn skills and theories in the context of real world, then construct their knowledge of a subject to be applied (Gay, Stefanone, Grace-Martin, & Hembrooke, 2001). This accords with the works of Brown, Collins and Duguid (1989) on situated cognition where they postulated that "knowledge is situated" and expounded on the necessity of cognitive apprenticeship and collaborative learning" where collective undertaking of authentic problem solving would give rise to new knowledge and insights. They attributed such a learning outcome to group dynamics and synergy. And interestingly, in theorising mobile learning, Sharples, Taylor and Vavoula (2005) reiterated that what essentially differentiates mobile learning from other types of learning is that "It is the learner that is mobile, rather than the technology" and they contended that a fundamental concern in understanding the essence of mobile learning is "to understand how people artfully engage with their surroundings to create impromptu sites of learning". Further, they highlighted the prerequisites of successful learning, point towards a social-constructivist approach on learning, where learning is conceived of "as an active process of building knowledge and skills through practice within a supportive community". Therefore, a working pedagogical framework for location-based collaborative learning should satisfy these two conditions: context- oriented and a collective enterprise.

One approach to support situated learning experiences in school is through the form of field trips (e.g., excursion, outdoor learning, etc). Previous research on field trip reveals that the educational effectiveness of a field trip is controlled by its structure, learning materials and the interaction with the environment (Orion & Hofstein 1994). Field trip visits often result in the students experiencing many phenomena and ideas that are new to them, where the individual is situated and ultimately have a strong influence on the ways in which knowledge is constructed (Anderson, Lucas et al. 2000). In the field trip, the environments cannot be pre-specified, therefore learning in the field trip including re-conceptualisation is created through the activity of learning. Knowledge is dynamically constructed by learners interacting with their surroundings (Sharples, Corlett, & Westmancott, 2002). This unique feature of location-based learning during field trips is expounded in Brown et al.'s (1989) concept of learning and cognition, where they posit that "situations might be said to co-produce knowledge through activity", hence learning and cognition are fundamentally situated.

Next, a social-constructivist approach also epitomizes the significance of collective co-construction of knowledge underpinning the nature of the learning process. Here, learners not only construct meaning on their own, but also with others to apply knowledge in real world context (Pena-Shaff & Nicholls 2004). Collaborative learning is thus, conceived of as active participation and interaction both with the environment and with others to negotiate meaning (Jonassen, Davison, Collins, Campbell, & Bannan Haag, 1995). Hence, the sense of place was emphasized in location-based learning since space not only refers to geographical locations, but also cognitive, psychologically affected space where social interactions place an important role (Gruenewald, 2003). And on the notion of technology-mediated mobile learning, mobile devices are mediating tools that allow people to capitalise on the situation in terms of the immediate physical space, while encouraging social communication and archiving in ways that enhance the learning context.

Methodology

Research Setting and Design Consideration

Premised upon a social constructivist framework on distributed cognitions (Salomon, 1993) and situated learning (Brown et al. 1989), the geography trail at Sentosa island in Singapore serves as a real world platform to acquaint students with situated collaborative learning; generating, sharing and affirming findings and solutions in problem-solving and inquiry-oriented activities. For the implementation of location-based geography trail, three main activity stations, named Yellow, Red, and Green, were identified along Sentosa's coastal areas spanning from Siloso beach to Palawan beach. To promote independent group decision-making, all instructions and procedures for activities were hosted on a web-based platform (see Figure 1), and facilitators, who are trained university students, were also present at all the activity stations. Facilitation was a built-in measure to assist students in their collaborative undertaking by providing them with task-oriented questions and necessary prompts that build on students' contributions to charter the course of their discussion. Each group was carrying Macbooks as mobile devices, and given a password to log onto the system and a set of coordinates to locate the stations with the help of Google maps. They were also able to upload their findings and collected artifacts onto these web pages created for each station.

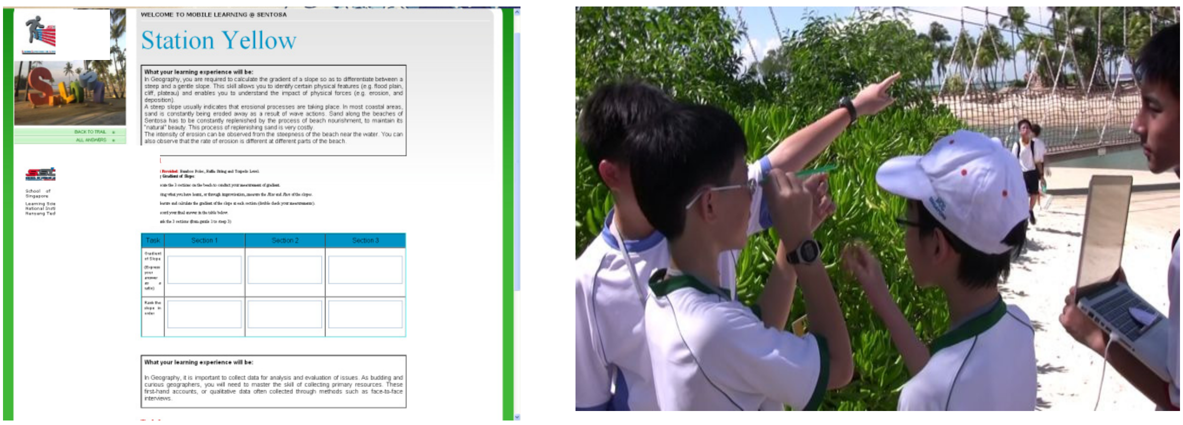


Figure 1. Web Platform (Left) and Students Measuring Height of Observation Tower Using Trigonometry (Right).

Table 1 presents an overview of the type of tasks designed for the mobile learning trail. The tasks were co-designed by the research team and collaborating teachers. Two key considerations drive the design and execution of the learning activities. Firstly, the learning activities should provide students with an authentic platform to apply their acquired geography skills and knowledge in a real world setting. Secondly, the activities ought to set the stage for in-situ collaborative learning, meaning collaborative meaning-making among students in the course of interaction with and within context. Hence, the on-site activities seek to maximize the presence of a *real world* platform, engaging students in meaningful knowledge creation and production where "the process of learning is informed by sense of place" (Lim & Barton, 2006, p.107). For instance, some of the key earning activities include measuring of gradient, identification of physical features in relation to the impact of physical forces of erosion and disposition, and the collection of qualitative data via face-to-face interviews with tourists.

In examining task design and implementation, we are particularly interested in the nature of task types in terms of its structuredness. In this study, well-structured tasks form *performative tasks* where learning paths to complete a task is rather fixed and procedural, leaving relatively little room for negotiation, judgment and conflict among group members. Measuring and calculating gradients of slopes (Task 1) is an example of well-structured tasks, as students can apply learned procedural skills to solve the task. On the other hand, we refer ill-structured tasks to *knowledge generative tasks* where the course of learning focuses on generating, communicating and co-constructing ideas. Design thinking task at the Green station (Task 7) falls into this category of knowledge generative tasks since the nature of tasks does not lead to one single answer or learning path; instead students need to propose workable solutions in the consideration of multiple dimensions. The nature of tasks can combine both performance and knowledge generation as the case of Task 2 where students need to generate interview questions and perform an interview with tourists.

Table 1: Overview of task design and characteristics.

Station	Task type	Description of tasks	Desired learning outcomes
Station Yellow	Performative	*Task 1:* Measure and calculate the gradient of the slope at 3 different sections of the beach and rank the slope from the gentlest to the steepest.	To understand the impact of physical forces such as erosion and deposition on the steepness of the beach.
	Performative and Knowledge generative	*Task 2:* Interview tourists to find out why they picked Sentosa as a holiday destination and what they think can be improved for Sentosa as a tourist attraction.	To collect qualitative data through primary resources such as face-to-face interviews for analysis and evaluation of issues.
Station Red	Performative	*Task 3:* Capture a picture along the coastal area and annotate five physical features: beach, island, observation towers, sea & suspension bridge.	To capture photo images and label its features as part of the process of data collection.
	Performative	*Task 4:* Calculate tower height using trigonometry	To estimate the height of both physical and human features & to relate the actual features seen on ground to the representation on topographical maps.
	Performative	*Task 5:* Identify, capture a picture of the ridge and annotate the physical feature.	To differentiate between physical features.
	Performative and Knowledge generative	*Task 6* Identify important industries near Sentosa and state their significance for the Sentosa establishment	To ask geographic questions, acquire and analyse geographic information
Station Green	Knowledge generative	*Task 7:* Design thinking with a focus on the beachfront area of the Sentosa island in terms of its attractions, accessibility and amenities. Identify a problem area and propose solutions, following the four fundamental steps of design thinking – brainstorm, share, categorise and solutioning.	To analyse, synthesize and evaluate real-life situations, in a systematic manner.

Participants

In this study, we employed a naturalistic case study method (Stake, 1995) to unpack the relationships among task design, discourse types, and facilitation. The location-based Geography trail was implemented at one Secondary school in Singapore in March 2010. To examine more closely how students co-construct ideas for each task type, our study focused on three groups of four Secondary One students each. The collaborating teachers initially expressed some concerns about gender and ability issues over technology-enhanced learning experiences that required complex problem solving skills. Hence, the selection of groups was randomly made in consideration of gender and academic ability (determined via a standard geography test). The three selected groups include two mixed-ability groups: Group 1 with all-girls and Group 2 with all-boys and one high-ability mixed gender group- Group 3.

Data Collection and Analysis

The conversation and interaction of group members for all activities on the learning trail were video- and audio-recorded and transcribed for analysis (73 pages in total). To analyse the students' interactions and knowledge construction process, all verbal data of the students was transcribed and analysed with respect to two key dimensions in the knowledge construction process, namely, *the epistemic* and *the social*. We adapted Pena-Shaff's framework (2009) for analyzing discourse where she subsumed the eight categories on knowledge construction proposed by Pena-Shaff and Nicholls (2004) into the epistemic and social dimensions surfaced in Weinberger and Fischer (2006). We found that while the frameworks were originally used to analyze online discourse, the coding categories are general enough to be applied in face-to-face discourse. Further, the frameworks were premised upon a social constructivism theoretical framework consistent with our theoretical stance on collective knowledge creation. Fischer et al. (2002 in Weinberger & Fischer, 2006) define the *epistemic dimension* as the on-task discourse where participants undertake knowledge construction tasks, and emphasis is given to the *content of the contribution*s of the participants. Of equal significance in the analysis of collaborative learning discourse is Weinberger and Fischer's (2006) proposition of the *social dimension* where they posit that the social modes of co-construction describes how learners response to and/ or build on the contributions of fellow participants in the course of solving a task. Table 2 features the revised coding category system for Knowledge Construction.

Table 2: Coding category system for knowledge construction.

Dimensions	Code Category and Description
Epistemic Dimension *Construction of Problem Space* (retell rather than interpret)	*Elementary Clarification (ECL):* Observing or studying a problem identifying its elements, and observing their linkages in order to come to a basic understanding
Epistemic Dimension *Construction of Problem and Conceptual Space and Relationships between both* (main task in problem-oriented learning environments)	*Question (QS):* Establishing relations between the problem and the conceptual space *Reply (RP)* *In-depth Clarification (ICL):* Analyzing and understanding a problem which sheds light on the values, beliefs, and assumptions which underlie the statement of the problem
Epistemic Dimension *Construction of Relations between Problem and Conceptual Space*	*Interpretation (IN)*
Some Epistemic Dimension	*Evaluation/Judgment (EV/JD)*
Statements not conductive to knowledge construction	Clarification without grounds; inappropriate use of theory/concepts, arguments with no ground, remarks on organization matters (*non-KC)*
Non-epistemic activities	Social comments not related to discussions (*non-EP)*
Social Dimension	*Externalisation (EX):* Discussions typically start with externalization/ Explain what they know *Elicitation (EL):* Receiving information from the learning partners learning partners as a resource *Quick consensus building (QCB)* *Integrated-oriented consensus building (ICB)* *Conflict-oriented consensus building (CCB)*

Coding Schema and Unit of Analysis

The group interactional discourse of completed tasks for all three experiment groups at the various activity stations was selected for analysis in this study. The number of tasks varies from station to station, ranging from minimum two tasks to maximum four tasks. As such, the corpus of transcribed verbal data at the different activity stations was segmentised first according to the different tasks at each station, and thereafter, according to semantic features such as topics, discussion threads and ideas. Chi (1997) argues that it is more meaningful using semantic boundaries to determine the unit of analysis since an idea may require a few sentences to put across; moreover, similar idea could be surfaced several times by team members who are more vocal. We found this particularly true of face-to-face collaboration where tracking the development of the interactions to trace evidences of collective knowledge construction is necessary. Hence, each segmentised unit of analysis in our study, depending on the discussion thread and topic, could contain a single or several categories. To achieve consistency and reliability, two more rounds of re-coding for all three participant groups were carried out by the first author. A second coder was also deployed, and an inter-rater reliability of 0.712 (Kappa coefficient) was achieved. In general, a Kappa value of 0.7 above is acceptable.

Findings and Discussion

A Comparison of the Occurrences of Discourse Types

The frequency of occurrences of statements for all task-discourse according to the category system is presented in Table 3 for all groups at all three activity stations. As mentioned earlier, only the conversational discourse of completed tasks is coded, thus for the tasks at yellow station, only findings of Group 1 and Group 3 are reflected and for tasks at red station, Group 3 completed only the first task out of the four tasks assigned. All the groups managed to complete the task at the green station.

Table 3: Frequency of occurrences of discourse types according to task-type for all three groups.

Task Structure	Performative									Performative + Knowledge Generative				Knowledge Generative		
	well-structured ill-structured															
	Task 3			Task 5		Task 4		Task 1		Task 2		Task 6		Task 7		
Groups	G1	G2	G3	G1	G2	G1	G2	G1	G3	G1	G3	G1	G2	G1	G2	G3
Epistemic dimension																
ECL	2	0	0	2	1	1	2	4	4	0	0	0	0	0	2	0
QS	1	1	0	1	1	0	3	0	2	1	1	1	2	3	5	1
RP	3	2	4	0	1	4	5	4	5	1	3	3	5	3	3	3
ICL	0	0	0	0	0	0	3	3	1	1	0	0	0	3	2	2
IN	0	0	0	0	0	1	1	2	2	2	2	1	2	8	7	4
EV/JD	0	0	0	0	0	2	1	3	1	3	2	0	0	5	6	3
non- KC	3	2	8	3	0	3	2	3	3	4	4	1	0	6	8	4
non-EP	2	0	7	0	1	4	1	2	0	0	0	0	0	1	0	2
Social dimension																
EX	2	3	6	2	2	6	5	12	6	1	3	3	1	3	6	3
EL	3	3	6	3	3	5	5	12	6	3	3	4	2	4	6	4
QCB	3	2	2	2	2	4	3	10	6	1	2	1	0	3	3	2
ICB	0	0	0	0	0	1	0	0	0	0	1	1	1	5	3	1
CCB	0	0	0	0	0	1	1	2	1	2	1	0	0	2	5	2

Note: ECL= Elementary Clarification, QS= Question, RP= Reply, ICL= In-depth Clarification, IN= Interpretation, EV/JD= Evaluation/Judgment, Non-KC= Statements not conductive to knowledge construction, Non-EP= Non-epistemic activities, EX= Externalisation, EL= Elicitation, QCB= Quick consensus building, ICB= Integrated-oriented consensus building, CCB= Conflict-oriented consensus building

A comparison of the occurrences of statements across all three stations reveals an unequal distribution of statement types. The bulk of the utterances during the task-discourse at yellow and red station for all three groups fall chiefly into basic clarification, reply, externalization, elicitation and quick consensus building categories. Clarification statements in the epistemic dimension remain mostly in the elementary phase except for Group 1 and Group 3 which manage to raise some in-depth discussion during Tasks 1 and 2 at the yellow station. For the red station, there are no occurrences of in-depth clarification, interpretation, judgment/evaluation, integrated-oriented and conflicted oriented statements for tasks 3 and 5. This unequal distribution of occurrences of statements becomes more apparent when one compares the task-discourse between the red and green stations for all three groups. The task-discourse at green station sees marked improvement in the occurrences of interpretation and judgment/evaluation across all three groups. Similarly in the social dimension, there are also reportedly fewer externalization and elicitation statements. Instead, we saw some instances of integrated-oriented and conflict-oriented statements. Statements not conductive to knowledge construction also witnessed a noticeable decline. Learners were more engaged in-depth discussions that require them to consider one another’s contributions, give comments and propose suggestions in the course of arriving at a consensus.

Impact of Task-design on Discourse Types

We argue that one reason to explain the above phenomenon lies in the *nature of task*. Tasks 3, 4 and 5 at the red station are chiefly application and procedural which require learners to transfer acquired geography knowledge and skills into real world situations. Hence, in the social dimension, statements made were pre-dominantly externalisation to verbalise what they already knew about a specific application or problem-solving process, and likewise, elicitation statements to seek affirmation of known procedures. Learners were also inclined towards quick consensus building to proceed with tasks since they were relatively familiar with task requirements and protocol. Likewise, we could also conclude that the task-type also accounts for the occurrences of basic clarification, question and reply statements in the epistemic dimension. Comparatively, the task on design thinking at green station (Task 7) allows more space for collective knowledge generation and construction. To visualize the difference between performative and knowledge generative tasks, we selected Tasks 3 and 7 and plotted the frequency of discourse types for each group in Figure 2. The column graphs show that the content of the contributions from the learners indeed show an increase in more in-depth clarification statements, evaluation, integrated-oriented and conflict oriented statements for Task 7 across all the three groups.

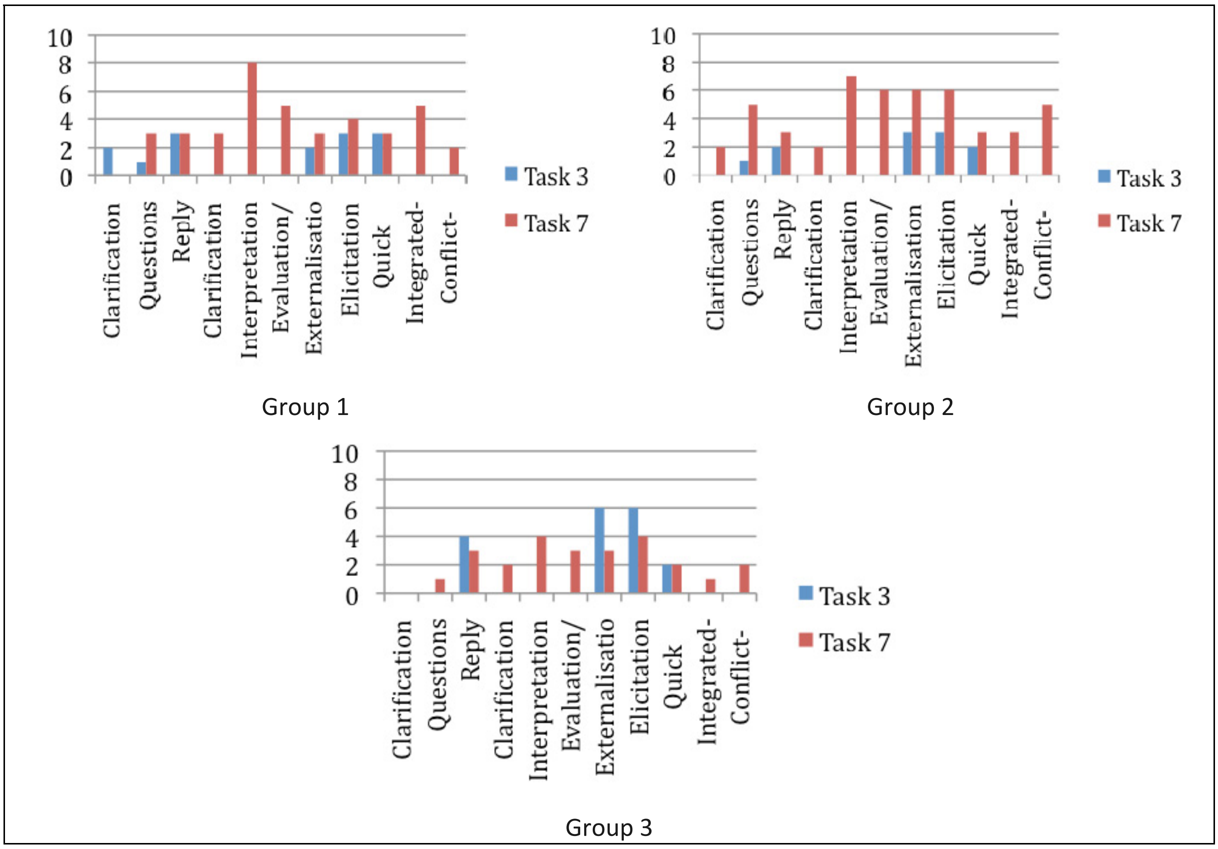


Figure 2. Discourse frequency charts between Performative Task (Task 3) & Knowledge Generative Task (Task 7).

Our findings and analysis further accentuated that different epistemic activities are also likely to dictate the learners' task-discourse. This inherently implies a close relationship between the types of epistemic activities and the social modes of co- constructing and advancing knowledge. On this note, Weinberger and Fischer (2006) posit that the varying social modes indicate both the degree and extent of "how learners work on the task and formulate arguments together (as opposed to individually)" (p.86). In a similar fashion, Salomon (1993 in Pena-Shaff & Nicholls 2004, p.244) postulates that new knowledge arises in the course of exchange, negotiation and transformation, and he contends that the "exchange of ideas and negotiation of meaning affects not only the individual's cognition but also the group's distributed cognitions". In the social process dimension, our case study also shows that pure application and/ or procedural tasks tend to yield more externalization, elicitation and quick consensus building, as surfaced in the conversational discourse for tasks at red and yellow stations. Discussion often remains at the elementary level of acknowledgement and affirmation of known procedures. And in the epistemic dimension, there are also significantly fewer evidences of a rigorous interaction in the sharing and improvement of ideas amongst learners to construct and advance knowledge.

Role of the Physical Environment in Relation to Task-design and Task Discourse Types

The essence of location-based collaborative learning is embodied in the constructivist approach on learning environments, where students are presented opportunities to think about the object and subject of study, construct meaning on their own and with others and to apply knowledge in real world context. A review of the task discourse for Group1 and Group 3 at yellow station Task 1 and Group 1 and Group 2 at red station Task 4 (see Table 3), shows that this task-type (though mainly skill-based and application) generated some in-depth clarification, interpretation and evaluation statements in the epistemic dimension, as well as, integrated and conflict-oriented consensus building in the social modes of co-construction. We attribute this phenomenon to the *affordances of in-situ learning*, where students are confronted with the real world platform to translate their acquired geography skills and knowledge into practice. Notwithstanding, the nature of tasks may not have changed drastically from the usual mundane procedural tasks which might have been accomplished in the four walls of a classroom, but the presence of the real physical environment certainly presented a different facet. The seemingly straight-forward application tasks such as the measuring and ranking of the gradient of slopes at the designated beach sections, and calculating tower height at the observation point saw unusual engagement in the task-discourse and greater collective knowledge construction. The interaction with the real environment presented some *unforeseen variables* where students found the application of known formulas was no longer as clear-cut, and the problem-solving process necessitated collective review of ideas and consensus. Students had to exercise more critical thinking in the course of finding and affirming solutions collaboratively.

Impact of Facilitation on Discourse Types in Relation to Task

Our findings surface another critical factor, which has significantly impacted the nature, intensity and scope of discussion, i.e. *facilitation*. It is evident from the relatively different occurrences of statements though the nature of task was similar, as in the case of Task 6 at the red station and the design-thinking task at the green station (see Table 3) that the nature and type of facilitation has a direct bearing on the group discourse. As mentioned above, the task-discourse at green station witnessed a visible increase of interpretation and judgment/evaluation across all three groups. Although this trend of discussion ought to have surfaced for Task 6 at the red station for Group 1 and Group 2 where the nature of the task also necessitated more collaborative critical thinking in the problem-solving process, there were only slightly fewer externalization and elicitation statements and sparse occurrence of interpretation and integrated-oriented consensus building statements. Taking a step further to re-trace the task- discourse for Group 1 and Group 2 at the red station, shows that where facilitator intervened frequently in the course of discussion and/ or posed highly structured questions, participants became less forth-coming with ideas. On the role of facilitation and its significance in collaborative learning environments, we concurred with Hmelo-Silver and Barrows (2008) that there exists a definitive relationship between "how the facilitator provided opportunities for knowledge-building discourse and how the learners accomplished collective knowledge building" (p.48). In other words, the type of questions posed will have a bearing on the *discourse moves* that inherently associate with effective collaborative knowledge building process and outcomes, in which Hmelo-Silver and Barrows (2008) spoke of *shared responsibility* of both the learners and the facilitator to move the discourse forward. Facilitation plays a significant role in culminating a productive discourse.

Conclusion

Overall findings provide accounts supporting the close relationship between task types and discourse types. When groups participated in the tasks that afford space for knowledge generation, the occurrence of interpretative and evaluative discourse appears to be high. On the other hand, when groups participated in performative tasks that are rather fixed and procedural, groups' discursive practices tend to be clustered around the sequence of question-answer and quick-consensus building. However, this is not to say that performative tasks are not desirable in collaborative learning practices. As we have presented in this paper, performative tasks do play important roles in supporting learners to internalize and externalize their knowledge and skills. Specifically in the context of location-based learning, applying knowledge and skills learned in classroom is not straightforward in authentic contexts due to unforeseen variables and complex interaction with the physical environment. Such complex situations of application are important learning opportunities for students to learn disciplinary problems through struggles, conflicts and even initial failures.

In relation to the conference theme on *connecting CSCL to policy and practices*, this study highlights the importance of paying close attention to the design configurations in mobile learning practices. In recent years, we have witnessed policy initiatives at the national and institutional levels to provide more affordable and accessible mobile devices to learners. In addition, there have been increasing interests about how to connect learning experiences in and out of school contexts through the mediation to mobile technology and applications. Field trips or outdoor learning supported with mobile technology is one form of situated learning that schools are exploring to provide students with rich, authentic experiences. We argue that whether such initiatives make any impact on collaborative learning practices in schools is closely associated with how learning tasks are designed and enacted to support authentic learning experiences. As this study shows, the dominance of performative tasks and/or highly structured facilitation is likely to reproduce the traditional discourse pattern like IRE in classrooms. Hence, we proposed that a three-prong approach (FAT) – Facilitation, Activities in-situ, and Technology – should be considered to support meaningful collaborative learning practices in mobile learning. To conclude, we believe that this study provides some insights about how the three intricate dimensions –task design, discursive practices, and facilitation – are unfolded in the situated context of collaborative learning. We argue that a balanced approach in task design is critical to make a transition from doing tasks to doing tasks *with understanding*.

References

Anderson, D., Lucas, K. B., Ginns, I. S., & Dierking, L. D. (2000). Development of knowledge about electricity and magnetism during a visit to ascience museum and related post-visit activities. *Science Education 84*(5), 658-679.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32–41.

Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *Journal of the Learning Sciences, 6*(3), 271-315.

Gay, G., Stefanone, M., Grace-Martin, M. & Hembrooke, H. (2001). The effects of wireless computing in collaborative learning environments. *International Journal of Human-Computer Interaction,* 13(2), 257-276.

Gruenewald, D. (2003). Foundations of place: A multidisciplinary framework for place-conscious education. *American Eeducational Research Journal, 40*(3), 619-654.

Hmelo-Silver, C. E. & Barrows, H. S. (2008). Facilitating collaborative knowledge building. *Cognition and Instruction*, *26*, 48–94.

Jonassen, D., Davison, M., Collins, M., Campbell, J., & Bannan Haag, B. (1995). Constructivism and computer mediated communication in distance education. *The American Journal of Distance Education, 9*(2), 7–26.

Lim, M. & Barton, A. (2006). Science learning and a sense of place in a urban middle school. *Cultural Studies of Science Education, 1*(1), 107-142.

Orion, N. & Hofstein, A. (1994). Factors that influence learning during a scientific field trip in a natural environment. *Journal of Research in Science Teaching, 31*(10), 1097-1119.

Pena-Shaff, J. B. & Nicholls, C. (2004). Analyzing student interactions and meaning construction in computer bulletin board discussions*. Computers & Education, 42*(3), 243- 265.

Pena-Shaff, J. (2009). Student patterns of interaction in asynchronous online discussions: Implications for teaching and research. In Méndez-Vilas, Solano Martín, Mesa González & Mesa González (Eds), *Research, reflections and innovations in integrating ICT in education*, 1. (pp. 440-445). Formatex, Badajoz, Spain.

Salomon, G. (1993). No distribution without individual's cognition: A dynamic interactional view. In G. Salomon (ed.) *Distributed cognitions - Psychological and educational considerations* (pp.111-138). Cambridge: Cambridge University Press.

Sharples, M., Corlett, D., & Westmancott, O. (2002). The design and implementation of a mobile learning resource. *Personal and Ubiquitous Computing, 6*(3), 220-234.

Sharples, M., Taylor, J., & Vavoula, G. (2005, October). *Towards a theory of mobile learning*. Paper presented at 4th World conference on mLearning, Cape Town, Africa.

Squire, K., & Klopfer, E. (2007). Augmented reality simulations on handheld computers. *Journal of the Learning Sciences, 16*(3), 371-413.

Stake, R. E. (1995). *The art of case study research*. Thousand Oaks, CA: Sage.

Weinberger, A. & Fischer, F (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education*, *46*, 71–95.

Acknowledgments

This research is supported by the FutureSchools@Singapore project under the National Research Foundation's (NRF) Interactive and Digital Media (IDM) in Education Research and Development (R&D) Programme. We thank SST teachers, students, and Yinjuan Shao for their contribution in this research.

Principle-Based Design for Collective Growth: from Knowledge-Sharing to Explanatory Knowledge-Building Discourse

Carol K.K. Chan, Ella L.F. Fu, The University of Hong Kong, Pokfulam, Hong Kong, SAR, China
Email: ckkchan@hku.hk; ellafu@hku.hk

Abstract: This study reports on classroom studies of knowledge building, supported with Knowledge Forum, conducted over two years with two teachers new to knowledge building and their Grade 4 students. This study examines how the two teachers engaged in increasingly sophisticated classroom design over time, and how their students moved towards more collective work and productive discourse. Obstacles and opportunities were identified as the researcher worked with the teachers to emphasize principle-based and collective design. Analyses indicated that, over time, students participated more in Knowledge Forum in terms of community awareness and connectedness, and moved more from knowledge-sharing to explanatory knowledge-building discourse. Analyses of students' posttest scores on science tests indicated that students' Knowledge Forum participation was positively correlated with domain understanding. Examining new teachers' shifts in classroom designs and their students' changes towards explanatory discourse may have implications for identifying the developmental trajectories of knowledge-building dynamics and practice.

Introduction

Helping students to engage in knowledge building is an important research theme in computer-supported collaborative learning and learning sciences. One particular model of knowledge building, also called knowledge creation, focuses on the creation and improvement of ideas and public knowledge in communities. When ideas are placed in the public, knowledge itself becomes an object of inquiry, and the community works collectively to make collective advances (Bereiter, 2002; Scardamalia & Bereiter, 2006). Knowledge Forum (KF), a computer-supported collaborative learning environment, is a knowledge-creation space that enables students to make their ideas public for progressive discourse and theory revision.

Over the last 20 years, major theoretical advances have been made with theory, pedagogy and technology of knowledge building (see special issue, Scardamalia, 2010). Research has shown that even young children are capable of engaging in collective cognitive responsibility to create new knowledge (Zhang, Scardamalia, Reeve & Messina, 2009). The knowledge building model has now been implemented by school systems in many countries, and there is an increasing need to understand the dynamics of knowledge building and how it can be implemented in a range of classrooms. Whereas attention has been given to how more experienced knowledge-building teachers engage in classroom design (Caswell & Bielaczyc, 2002), less is known about students' knowledge-building trajectories and knowledge-building practice among novice teachers. The goal of this study is to characterize knowledge-building discourse, in order to identify developmental trajectories and to explore the designs that facilitate teachers to make these changes. Specifically, this study examines teacher and student growth trajectories by following, over the course of two years, two Hong Kong teachers who teach the same curriculum topics and are both new to knowledge building.

Developmental Patterns of Knowledge-Building Dynamics and Discourse

Since the 1990s, considerable research has been conducted for characterizing knowledge-building processes and dynamics. Different patterns of knowledge building dynamics have been identified, for example, referent-based versus problem-centered knowledge (Oshima, Scardamalia & Bereiter, 1996); fact-based versus explanatory-based inquiry (Hakkarainen, 2003); and depth of questions and explanations (Lee, Chan & van Aalst, 2006) More recently, van Aalst (2009) has identified three kinds of knowledge building discourse – knowledge sharing, knowledge construction, and knowledge creation – associated with the emergence of new knowledge and social dynamics. Such kinds of analyses, rather than merely coding and counting CSCL discourse categories, imply increasingly sophisticated patterns of processes and dynamics in knowledge-building discourse. The question is, can these patterns of discourse be identified to characterize developmental trajectories of productive knowledge building?

Analyzing knowledge-building dynamics and trajectories is useful, not only for characterizing the nature of knowledge building, but also for examining changes in teacher and student knowledge building practices. Hakkarainen (2003) report on changing from fact-based to explanatory-based inquiry over a three year period; and Zhang *et al*'s (2009) three-year design-study into changing fixed, interactive group dynamics to opportunistic group dynamics with increased community awareness and community connectedness, for example – have specifically examined changes over time. These patterns may suggest pointers for identifying and developing more sophisticated processes and practices. Nevertheless, most of these studies have involved comparatively sophisticated teachers, rather than investigating knowledge-building dynamics trajectories by examining novice teachers' growth towards adopting knowledge-building innovation and tracking changes in their students' discourse. The scheme of knowledge-building discourse (van Aalst, 2009) was identified among Canadian high-school students; it would be useful to examine further possible developmental patterns and to investigate whether and how elementary school children change from knowledge-sharing to knowledge-construction and to knowledge-creation discourse.

From Activity-Based to Principle-Based Design

Questions also exist as to how new teachers can adopt more sophisticated knowledge-building practice to facilitate more productive knowledge-building discourse. A major challenge for teachers working on inquiry-based and knowledge creation model is to shift from tasks and activities to principle-based understanding for classroom innovation. The traditional classroom model is one of highly-structured initiate-response-evaluation (IRE) patterns, with a focus on individuals, so how can teachers move towards designs for collective cognitive responsibility? Scardamalia and Bereiter (2006) contrast principle- and procedural-based approaches: A key idea in principle-based innovation is that teachers need to go beyond activities and view principles as objects of inquiry; knowledge-building principles can point to and scaffold collective designs(Scardamalia, 2002). There is some evidence that principle-based understanding sustains teacher growth in school innovation (Zhang, Hong, Scardamalia, Reeve, & Messina, in press). Research has also shown that teachers in a knowledge-building network vary in how they emphasize principles versus activities, and that the students of those inclined towards principle-based understanding have more sophisticated views of collaboration (Chan, in press). Accordingly, this study examines the idea that as new teachers move towards principle-based collective designs, they are better able to bring about changes in their students' productive discourse.

To summarize, the research objectives for this study are to: (1) examine teachers' classroom designs and to investigate how they change from activity- towards principle-based designs over a two year period; (2) investigate how teacher change in classroom designs parallel their students' change towards increased community connectedness and explanatory knowledge-building discourse; and (3) examine whether student engagement in knowledge building activity is related to their domain understanding.

Methods and Design

Participants

The participants were two elementary-school teachers (hereafter called TC and TT) and their Grade 4 students, who were examined over two school years (2008-09 and 2009-2010). TC and TT had 9 and 5 years of experience respectively, when they joined the project. Both taught in a public school located in a low income area of Hong Kong with children of low-average academic levels. In Year 1, TC's and TT's classes had 40 and 33 students respectively; in Year 2, the numbers were 39 and 29. The two teachers implemented the knowledge-building approach in both semesters of both years; the selected topics were "Plants" and "Human Body". This design made it possible to track how these new teachers made change and as reflected in their student participation. To control for differences in student background and improvement due to a strong cohort, school information was collected. Students in Year 1 and Year 2 were similar in terms of School Attainment Test scores (similar to standardized test scores), with Year 2 students in both classes being generally weaker.

Designing and Implementing Knowledge Building

Year One Implementation – Initial Work

While the two teachers both subscribed to student-centered learning, both emphasized steps and procedures: At the start, they listed their strategies: Students would (a) read materials in library; (b) watch documentary films to stimulate interest; (c) ask 'wh' questions about the topic; (d) select the three most popular questions for KF work; (e) learn KF functions; (f) participate in student discussions using KF; and (g) produce concept maps and a learning diary. TC began with implementation – he was concerned that he could not deal with students raising too many questions, thus only three questions were chosen for KF discussion. Figure 1a shows an interesting pattern that is characteristic of teachers and students new to knowledge building – although many notes were made, most were short and fragmented, and formed a kind of 'spider-web' pattern around the centralized node as students wrote responses to the parent question.

Ongoing evaluation indicated that, while students were relatively engaged and that many notes were posted, the responses were superficial and repetitive. We worked with the two teachers to identify areas needing refinement. Fig 1a shows that students viewed discourse as a question-answer pattern, analogous to the teacher asking a question that had to be answered. Moreover, they did not see that the discourse is intended to build on and extend others' ideas for community advances, and kept giving the same responses, mostly short answers with no explanation. Students typically just answered the question or clicked on a random note. TT and

TC worked with each other - While TT also used three questions and responses were still fragmented and descriptive, the spider-web pattern seen in TC's class did not exist in his. TT and TC were sufficiently impressed with what their low-achieving students could do using KF, that they continued the project with the next semester's topic, the Human Body. After an initial phase of question-answer responses, students became more engaged in posing questions and building on each others' responses. Discourse was still superficial but improving to more divergent pattern (Fig 1b) -- students tended to ask questions about difficult terminology (what is X or Y?) without linking them to the problem at hand.

Year Two Implementation – Towards Principle-Based Design

Throughout Years 1 and 2, TT and TC worked with the researchers and other peer teachers in the context of a knowledge-building teacher network (Chan, in press). Teachers were encouraged to view classroom design problems as objects of inquiry and they also employed technology in ways that are connected with the principles. Focus was placed on principle-based design emphasizing individual to collectivist stances. As TT and TC worked on the same topics with parallel classes, they helped scaffold each others' advances. The change process was gradual and emergent; classroom designs of Year 2 informed with principles are described:

Epistemic agency. TC and TT no longer restricted the number of questions posed by their students, and became more comfortable letting students ask the questions they needed to have answered. TT noted that students would give deeper responses when they felt they were asking questions to which they genuinely wanted answers. While their goal is to help students develop collective agency, it was not an exercise in discovery learning; they worked with their students to connect their wonderment with big ideas in the domain.

Community knowledge. The teachers addressed the problem of repetition by using the analogy that people in a conversation do not keep repeating what others have already said; rather, they listen to what has been going on and move forward the conversation. The teachers helped students to change gradually from the predominant question-answer patterns to ones that involve interaction, contribution, and discourse -- A divergent and distributed pattern now emerged. While TT and TC had asked students to work in groups in Year 1, later they created an overall view and used rise-above notes to develop community knowledge (Figure 1c).

Improvable Ideas. To help students move past shallow knowledge-sharing, a major change was introduced in Year 2 – the use of "knowledge-building talks" in which students' emerging questions were reflected upon. TC worked with students on what he called 'cutting-edge' recommended notes. In the past, some important ideas might have gotten lost; and students shied away from difficult questions; now the teachers encouraged students to help each others to improve their ideas.

Constructive use of information. TC and TT provided their students with relevant information to encourage them to move beyond chit-chatting and sharing opinion. Both teachers were working on the problem in that some students simply copied the information from the web into their KF notes without interpretation.

Embedded assessment. While the teachers were concerned with assessment, in Y1, they asked students to reflect on their understanding using concept map in paper and pencil. Over time in Year 2, both teachers were better able to use collective assessment with reference notes on Knowledge Forum..

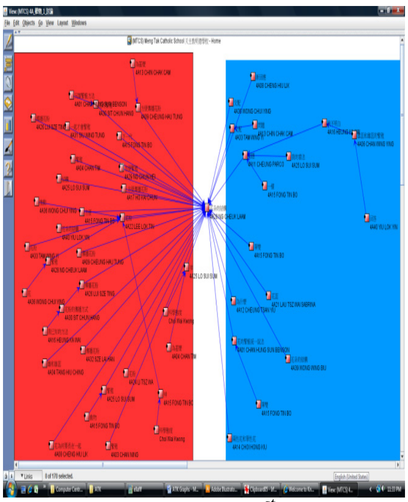


Figure 1a. KF View in 1st Yr TC.

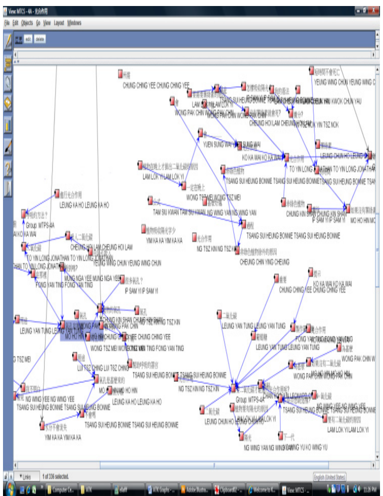


Figure 1b. KF View in 2nd Yr TC.

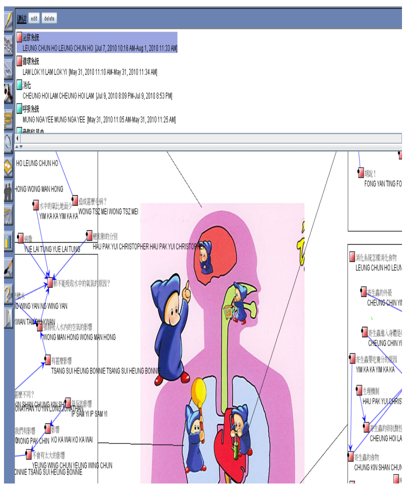


Figure 1c. KF View in 2nd Yr TC.

Results

Changes in Students' Knowledge Forum Participation over Two Years

Community Awareness and Connectedness (Applets –Note-Reading and Build-On Density)

Data were collected to examine whether the two teachers had changed over time, based on their students' participation and collaboration patterns on Knowledge Forum. The mean number of notes written were 8.7 to 16.7 for TC and 4.5 to 19.5 for TT' classes in Yr 1 and 2. We employed two Knowledge Forum assessment Applet indices – "Note-Reading Density" and "Build-on Density" to gauge student progress in the knowledge building community; the indices are similar to Social Network Analyses, but simpler to use. Density is calculated as the ratio between the actual number of links and the maximum possible number of links. Zhang et al. (in press) note that the extent to which students read others' notes reflects "community awareness" and that building on reflects "community connectedness"; both are used here to indicate whether students are moving towards more community cohesion. There were changes in Read-Density (TC: 84.5%--100%; TT: 79.3%—93.65%) and Build-on Density (TC: 29.2% –32.7%; TT 18.6%--43.9%) over the two year period. Figures 2 and 3 show the changing patterns using Applet visualization patterns for build-on notes. These results indicate that students of both teachers had more connected interactions on KF participation comparing Year 1 and Year 2.

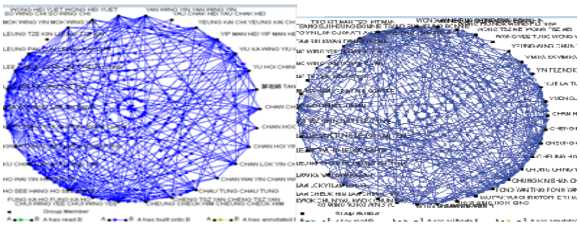


Figure 2. Note-Build-on Density in Years 1 and 2 (TC).

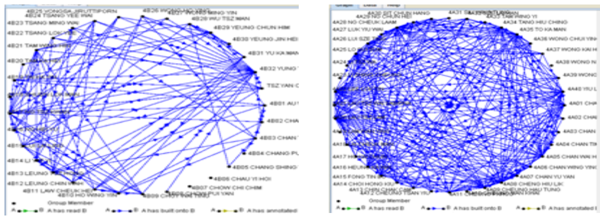


Figure 3. Note-Build-on Density in Year 1 and 2 (TT).

Participation and Collaboration (Analytic Toolkit Indices)

Student participation on KF was obtained using server log information via Analytic Toolkit, a software with Knowledge Forum.. Several indices used in previous studies were generated including "write", "revision", "scaffolds", "keywords", "read", and "link" (Lee et al., 2006). Figs 4 & 5 show changes in student participation over the two years for the ATK indices. To provide a more coherent analysis, the ATK participation indices were factor analyzed, and two factors were obtained called "Productivity" (write, read, revise, scaffolds) and "Collaboration" (linked, keywords), consistent with findings in other studies (Lee et al., 2006). Statistical analysis indicates marginally significant differences between Years 1 and 2 for ATK Productivity, F =3.34, p <.08; and significant differences for ATK Collaboration, F = 4.17, p <.05 with higher ATK scores in Year 2. When the two classes were examined separately, there were differences for TT's class but not TC's classes. Taken together, these two sets of quantitative indices (Applets + ATK) suggest that there was more participation, collaboration and connectedness among student cohorts for the teachers over the two years.

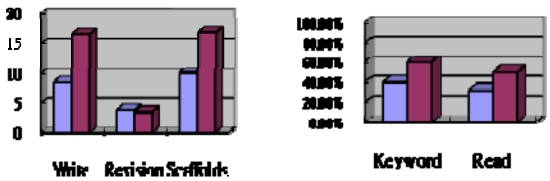


Figure 4. ATK Forum Participation Indices (TC).

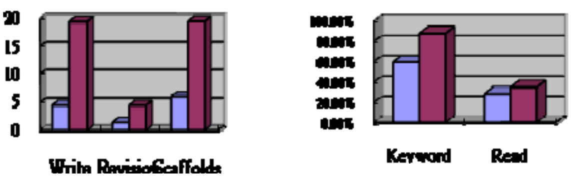


Figure 5. ATK Forum Participation Indices (TT).

Changes in Discourse Over Two Year

Characterizing and Analyzing Knowledge-Building Discourse

Students' forum writing was analyzed for discourse quality by adapting Zhang et al' notion of conceptual threads (2007) grouping forum notes into different discussion/inquiry threads. These discourse threads rather than individual notes or adjacent pairs were the unit of analysis for tracking changes in knowledge quality. Discussion threads were then coded into four levels using the idea of different types of knowledge-building discourse (van Aalst, 2009):

- *Fragmented discussion.* Short discussion threads with fragmented responses; an example is the pattern of short answers and questions as described in Figure 1.
- *Knowledge-sharing discourse.* There are two types: short threads in which a student asks a question and someone responds often using information copied off of the internet. The second involves conversational interaction in which opinions or information is shared without supporting evidence or explanation. Table 1 shows an example in which a student asks what might happen if a person had only one lung. Several

students respond with either opinions or general ideas (#3-6). There was interaction but no attempt to formulate the problem to understand the respiratory system. The discourse then degenerates into conversations about lung disease and that thinner people are weaker and more susceptible to lung diseases (#11-12). Although students are interacting, they are primarily exchanging opinions.

Table 1: An example of a knowledge-sharing thread.

Student	Discourse Moves	Excerpts from the inquiry thread
1. HSH	Asks a question	I need to understand We have a pair of lungs. If we only have one side of the lung, can we still breathe? Why?
2. CCY	shares opinion	My theory If we only have one side of the lung, we can still breathe, because the only reason for us to have a pair of lungs is to breathe smoothly. If we only have one side of the lung, we cannot breathe smoothly.
3. LYH	shares opinion	My theory Lung helps us to breathe. We can breathe with one side of the lung. We will certainly die, if we do not have a pair of lungs.
4. CT	shares opinion	My theory I guess we cannot breathe with one side of the lung.
5. CWY	shares opinion	My theory Can breathe.
6. YMH	shares opinion	My theory Yes, but it would be much harder. The job to breathe should be done by a pair of lungs, but now it is done by one side of the lung only.
7. WHY	share opinion	My theory One side of the lung is also a respiratory system, though breathing is pretty hard.
8. CWY	Asks a question	I need to understand why do we not die with one side of the lung?
9. LYK	digression	I need to understand What will happen if there is a hole in the lungs?
10. CCY	shares opinion	My theory It is called burst lungs. It makes breathing very difficult. It usually happens in people who are thin and tall.
11. CWY	Asks a question	I need to understand Why people who are thin and tall often have burst lungs?
12. YHC	shares opinion	My theory Because those thin and tall people are weak.

Notes: Students wrote in Chinese on KF; these excerpts are translated for illustration. Many discussion threads are long, and only selected notes are included here to highlight the key features.

- *Knowledge-construction discourse*. Attempts were made by students to construct understanding using questions and explanations. Table 2 shows an example in which a student asks why plants release carbon dioxide at night (#1). Another student responds that plants release carbon dioxide and absorb oxygen at night and the other way round in the morning, which leads to a discussion of the role of CO2 in photosynthesis (#3-5). There are good efforts to build on others' ideas with questions and explanation.

Table 2: An example of a knowledge-construction thread.

Student	Discourse Moves	Excerpts from the inquiry thread
1.WTM	asks a question	I need to understand Must plants release carbon dioxide in the evening? Why?
2.LLY	elaborates and includes information	My theory Plants must release carbon dioxide at night because, in the morning, plants absorb carbon dioxide and release oxygen. In the evening, plants absorb oxygen and release carbon dioxide.
3.MT	asks a question	I need to understand Why do plants need carbon dioxide to have photosynthesis?
4.CHJ	explanation	My theory Because carbon dioxide is a raw material for nutrition production…
5. LLY	Uptake; elaborates & explains	My theory Plants need carbon dioxide to have photosynthesis because some procedures of photosynthesis require the combination of the carbon dioxide and glucose for oxygen…
6.KWY	explanation	My theory If there is no carbon dioxide, plants cannot have photosynthesis to produce nutrition. Plants will die and not have reproduction.

- *Emerging knowledge-creation discourse.* Table 3 shows an example with sophisticated explanatory and emergent discourse. The thread begins with a question about how non-green plants produce food, as they may not have chlorophyll. The initial response is a misconception, but allows others to seek clarification and leads to revision for better understanding. One student deepens the inquiry, asking how red leaves can have photosynthesis (#5), to which another student discusses how to test the theory that red leaves can make food through a laboratory experiment (#6). These responses, from Grade 4 children, reflect sophisticated discourse -- puzzlement are raised that spawn questions; theory testing through experiments and theory building using new information and explanation (#7-8) This emerging knowledge creation thread is characterized by question-explanation, emerging questions and theory testing.

Table 3: An example of an emerging knowledge-creation thread.

Student	Discourse Moves	Excerpts from the KF note
1. MT	asks a question	I need to understand How do non-green plants produce chlorophyll and nutrition?
2. YLT	poses idea thus misconception can be inquired further	My theory Non-green plants cannot have photosynthesis; therefore, they are parasite on plants that have chlorophyll. They absorb nutrition from other plants.
3. TSH	indicates puzzlement & asks for explanation	I need to understand How can non-green plants be parasite on green plants, and how can they absorb nutrition from other plants? (What is the process of such absorption?)
4. LLY	Provides explanation	My theory In fact, similar to green-plants, non-green plants contain chlorophyll. Non-green plants do not need to produce chlorophyll, and they produce nutrition the same way as other plants.
5. KCY	asks a question	I need to understand How do red leaves have photosynthesis? Why?
6. MHH	hypothesizes and proposes an experiment	My theory Red leaves also have chlorophyll, even though such chlorophyll is red in color instead of green in color. We can conduct the same experiment shown in the textbook… Put red leaves in a test tube, and then pour iodine on the leave… We can test whether the red leaf can have photosynthesis or not?
7.KCY	Clarifying question	I need to understand Is it call "red" chlorophyll?
8. LKH	Uses new information to construct explanation; deepening inquiry (my theory--new information--my theory)	My theory … the major pigment in chlorophyll is green in color. When plants have a lot of chlorophyll, they are green plants. New Information …For example, red algae contain "algae bile pigment"; brown algae contain a "black algae yellow substance."… My theory Those color [crowd out] the green color in chlorophyll. That is why we see non-green plants. The non-green plants use those … colors to absorb the light and move the energy to chlorophyll? Such other colors are indirectly involved in the process of photosynthesis…

Note: The Chinese translation of chlorophyll includes the word 'green', thus students were puzzled about whether non-green plants have chlorophyll and how they can make food.

Changes in Quality of Discourse and More Sustained Inquiry

Student discourse in both years and with both teachers was grouped into discussion threads and coded into the four levels (Table 4) indicating an improvement in discourse patterns. Specifically, the number show that knowledge-sharing patterns decreased over time, while knowledge-construction threads and emergent knowledge-creation threads increased. Ongoing analyses were conducted to establish inter-rater reliability for the coding of these threads. Since knowledge building involves sustained inquiry, to provide a fuller picture, the threads were plotted to illustrate the extent to which students were continuing with their inquiry over time versus short threads for question-answer. Figures 6 and 7 show the discourse threads for the topic 'plants' for TT and TC (Yr 1 & 2). The plots of the discussion threads indicate some distinctive differences – Comparing Years 1 and 2, students wrote more threads, and these threads lasted for more days suggesting that students of these teachers were more engaged and sustained in their knowledge-building inquiry.

Relations between Participation on Knowledge Forum and Domain Understanding

The third research question examines whether student engagement with knowledge building measured by their participation on Knowledge Forum is related to their domain understanding, We collected data from (a) a domain test that consists of two questions (what have you learned? what would you like to find out?) and (b) examination scores in general science as students' regular assessment practice in schools. We followed the approach employed in other studies (Lee et al., 2006) and coded student responses using a 5-point scale ranging from fragmented to scientific responses. The four scores from the two units (Plants and Human Body) were combined, yielding mean domain test scores of 2.5 (TC) and 2.8 (TT). Analyses were conducted to examine if student participation on Knowledge Forum, as measured by ATK indices, was related to domain understanding. As mentioned above, factor analyses of the indices generated two factors – ATK "Productivity" and ATK "Collaboration". Correlation analyses indicated that students' domain test scores were correlated with ATK Collaboration (r = .26, , *p* <. 04); and school exam scores were correlated with ATK Productivity (r = .28, *p* <.03). Separate analyses showed stronger patterns in TC's class. Overall in both classes, students' participation indices on Knowledge Forum were significantly related to their domain understanding.

Table 4: Changes in discourse patterns for student cohorts in two years.

	TC (4A)				TT (4B)			
	2008-09		2009-10		2008-09		2009-10	
	No.	%	No.	%	No.	%	No.	%
Fragmented	1	4.2	1	2.78	7	50	0	0
Knowledge Sharing	20	83.3	25	69.4	5	35.7	23	69.7
Knowledge Construction	3	12.5	6	16.7	1	7.1	7	21.2
Emerging Knowledge Creation	0	0	4	11.1	1	7.1	10	9.09
Total	24	100	36	100	14	100	33	100

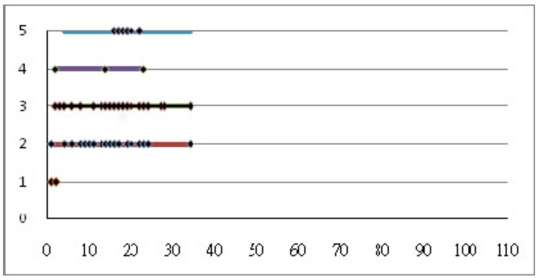


Figure 6a. Short Discourse Threads (Year 1 TC).

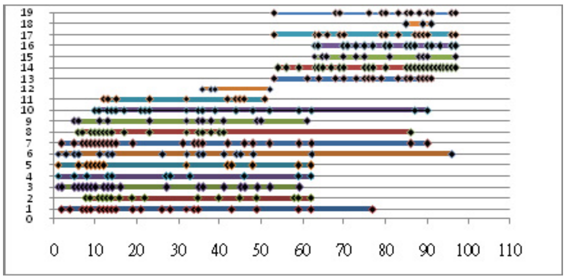


Figure 6b. Sustained Discourse Threads (Year 2 TC).

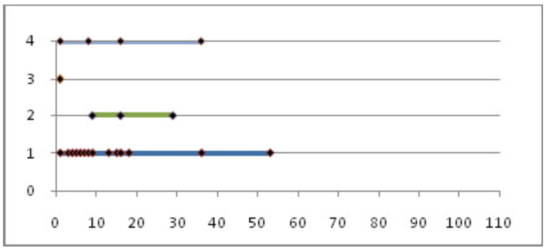


Figure 7a. Short Discourse Threads (Year 1 TT).

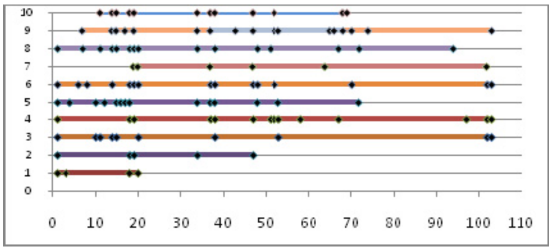


Figure 7b. Sustained Discourse Threads (Year 2 TT).

Discussion

Despite the enthusiasm the knowledge-building model inspires, its emphasis on collective work is difficult for teachers and students alike. Whereas many knowledge-building studies report on experienced teachers, this paper examines the challenges facing new teachers as they gradually make changes towards more principle-based designs, and how their students develop more cohesive interactions and productive discourse.

This study has examined the characterization and developmental trajectory of knowledge-building discourse in the context of how two teachers move towards more sophisticated knowledge-building design over time. The first question examined the teachers' change from activity-based to principle-based and from individual to collective designs. A key theme is that the two teachers, working with the researchers and the teacher network, were engaged in a design-based process, examining student problems and advances while making needed refinements. The teachers' initial design restricted the number of questions students posed; the results showed spider-web patterns that reflected students' beliefs about how questions are to be answered and other related problems, such as repetitive responses and lack of explanation. Examining these debilitating patterns may help new teachers to understand key principles of knowledge building, and to emphasize epistemic agency and community knowledge. There are various ways in which the teachers worked with students, employing key principles, such as helping students to identify productive areas of inquiry for improvable ideas. Figure 1 shows the gradual progression of student work on KF – from spider web (Fig. 1a), to divergent responses (Fig. 1b), to an integrated view (Fig. 1c); and there were changes from using a paper-pencil concept-map to KF reflective assessment with reference notes reflecting attempts towards a meta-discourse. At various phases, the teachers were concerned with continual refinements and how their students' work could be improved. The two new teachers worked closely supporting each other; their patterns of results suggest both made progress over the years. Examining how new teachers adapt different designs and grappling with principles may have implications for knowledge-building design and professional development.

Another key theme of this paper characterizes the nature and development of the knowledge-building discourse through examining knowledge-building practice at different times. The second question examined knowledge building participation and discourse among the children over the two-year period. Quantitative and qualitative analyses show that children in Year 2, compared to Year 1, participated more actively on Knowledge Forum, with increased community connectedness, and they were more engaged in productive discourse. These results suggest that changes in teacher practice using principle-based design might have helped to bring about more community cohesion and sustained discourse in the class community. The third question indicating the relation between Knowledge Forum engagement and domain understanding is reassuring to the teachers; the results are consistent with other studies conducted in Hong Kong classrooms (see Lee et al., 2006).

Analyses indicate discourse of school children in Hong Kong can be differentiated into different types including knowledge sharing, knowledge construction, and emerging knowledge creation (van Aalst, 2009). For knowledge sharing, children are interacting but they are involved with sharing opinion with little explanation or evidence. Knowledge construction suggests their efforts to develop some understanding; question-explanation is important discourse moves. For emerging knowledge creation, it seems that even elementary school students can engage in scientific discourse, reflecting efforts towards collective inquiry and theory building and the notion that ideas and theory can be tested and revised through experimentation. The changes among teachers and students over the two years suggest possible developmental progression that has theoretical and design implications. The different kinds of discourse threads can be employed by teachers for examining students' work and scaffolding students towards more productive discourse. Further investigation can be conducted to investigate how teachers shift from procedures to principles as they scaffold their students to change from knowledge-sharing towards explanatory-based knowledge-creation discourse.

References

Bereiter, C. (2002). *Education and mind in the knowledge age*. Mahwah, NJ: LEA.

Caswell, B., & Bielaczyc, K. (2002). Knowledge Forum: Altering the relationship between students and scientific knowledge. *Education, Communication & Information, 1* (3), 281-305.

Chan, C. K. K. (in press). CSCL theory-research-practice synergy: Implementing knowledge building in Hong Kong classrooms. *International Journal of Computer-Supported Collaborative Learning*.

Hakkarainen, K., Lipponen, L., & Jarvela, S. (2002). Epistemology of inquiry and computer-supported collaborative learning. In T. Koschmann, R. Hall & N. Miyake (Eds.), *CSCL 2: Carrying forward the conversation* (pp. 11-41). Mahwah, NJ: Lawrence Erlbaum Associates.

Hewitt, J. (2002). From a focus on task to a focus on understanding: The cultural transformation of a Toronto classroom. In T. Koschmann, R. Hall & N. Miyake (Eds.), *CSCL2: Carrying Forward the Conversation* (pp. 11-41). Mahwah, NJ: Lawrence Erlbaum Associates.

Lee, E. Y. C., Chan, C. K. K., & van Aalst, J. (2006). Students assessing their own collaborative knowledge building. *International Journal of Computer-Supported Collaborative Learning, 1* (1), 57-87.

Oshima, J., Scardamalia, M., & Bereiter, C. (1996). Collaborative learning processes associated with high and low conceptual progress. *Instructional Science, 24*, 125-155.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal education in a knowledge society* (pp. 67-98). Chicago, IL: Open Court.

Scardamalia, M. (Ed.) (2010). Knowledge building. [Special Issue]. *Canadian Journal of Learning and Technology. 36* (1).

Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 97-115). New York, NY: Cambridge University Press.

van Aalst, J. (2009). Distinguishing knowledge-sharing, knowledge-construction, and knowledge-creation discourses. *Computer-Supported Collaborative Learning, 4*, 259-287.

van Aalst, J. & Chan, C.K.K. (2007). Student-directed assessment of knowledge building using electronic portfolios. *Journal of the Learning Sciences*, *16* (2), 175-220.

Zhang, J., Hong, H. Y., Scardamalia, M., Toe, C. & Morley, E. (in press). Sustaining knowledge building as a principle-based innovation at an elementary school. *Journal of the Learning Sciences.*

Zhang, J., Scardamalia, M., Reeve, R., & Messina, R. (2009). Designs for collective cognitive responsibility in knowledge building communities. *Journal of the Learning Sciences*, *18*, 7-44.

Does Contributing to a Knowledge Building Dialogue Lead to Individual Advancement of Knowledge?

Maria Chuy, Institute for Knowledge Innovation and Technology, OISE/University of Toronto, 252 Bloor Street West, Toronto, ON M5S 1V6, Canada, maria.chuy@gmail.com
Jianwei Zhang, University at Albany, 1400 Washington Ave, Albany, NY 12222, USA, jzhang1@albany.edu
Monica Resendes, Marlene Scardamalia, Carl Bereiter, Institute for Knowledge Innovation and Technology, OISE/University of Toronto, 252 Bloor Street West, Toronto, ON M5S 1V6, Canada,
Email: monicaresendes@gmail.com, marlene.scardamalia@utoronto.ca, carl.bereiter@gmail.com

Abstract: The goal of this research is to examine the extent to which contributing to a Knowledge Building online dialogue can predict individual advancement of knowledge for Grade 4 students. Based on the procedures of grounded theory (Glaser & Strauss, 1967), we examined notes that grade 4 students entered into an online discourse environment (Knowledge Forum) and developed the following empirically grounded list of contributor roles: asking thought-provoking questions, theorizing, experimenting, working with evidence, creating syntheses and analogies, and supporting discussion. Regression analysis was performed to examine whether any of these roles can predict various measures of individual knowledge advancement. The analysis revealed two significant predictors: theorizing and working with evidence. Theorizing accounts for variation in scientificness of students' ideas as well as knowledge test scores; working with evidence predicts epistemic complexity of students' portfolios. These results inform next steps in creating classroom interventions and technology tools for collaborative knowledge creation.

Introduction

Knowledge creation and innovation have risen to high prominence in the economic sphere (David & Foray, 2003) and for societal problem solving (Homer-Dixon, 2006). From this perspective, capacity for knowledge creation represents a major challenge for education (Bereiter & Scardamalia, 2006; Carnevale & Desrochers, 2003: Scardamalia & Bereiter, 2007). Among educational approaches that can claim special relevance in this respect (e.g., project- and problem-based learning, "21st Century Skills" programs), "Knowledge Building" stands out as most directly addressing the need for knowledge-creating talent. Defined as the deliberate creation and improvement of knowledge that has value for a community (Scardamalia & Bereiter, 2003), Knowledge Building is synonymous with knowledge creation. Individual knowledge workers, often in teams, make diverse contributions to the collective knowledge in a domain through efforts that preserve and elaborate the current paradigm, make incremental moves in existing directions, or redirect the field into a new direction or paradigm (Sternberg, 2003). Thus, Knowledge Building challenges students to be not only learners but active contributors to group efforts to produce epistemic artifacts—explanations, problem formulations, proofs, experimental methods, "state of the art" reviews, and the like. These epistemic artifacts are collectively improved through a wide variety of individual student contributions, ranging from drawing useful diagrams to suggesting pertinent analogies and theories. Being complimentary, these contributions form a highly interactive "self-organizing" system (Resnick, 1996), moving knowledge-creating dialogue forward without central authority. Both parts symmetrically benefit from this interaction: the community, as a whole, and individual contributors. However, individual benefits might fluctuate, depending on the degree of the individual's participation and his/her contributor roles in the community's dialogue. That is the focus of the present research—examining the extent to which contributing to a Knowledge Building online dialogue can predict individual advancement of knowledge in elementary school students.

A number of collaborative knowledge creating approaches make use of assigned roles. Edward de Bono (1985) distinguishes six different thinking strategies symbolized by six colored hats: a black hat is associated with critical thinking, a green hat with idea generation, a white hat with providing information and facts, etc. Leng, Lai, and Law (2008) elaborated a taxonomy of ways of contributing on the basis of levels of cognitive complexity. Hogan (1999), observing student interactions in science discourse, identified eight naturally occurring roles similar to de Bono's "six hats:" promoter of reflection, contributor of content, creative model builder, mediator of group interactions and ideas, etc. The current research aims to develop an empirically grounded list of contributor roles of a finer grain than the above, and to examine the extent to which these roles can predict individual knowledge gain in the context of Knowledge Building. Thus, analyses focus on: (a) creating a systematic inventory of contributor roles, based on analysis of student online discussions; (b) identifying measures of individual knowledge advancement; and (c) studying the relationships between these two components.

Method

Knowledge Building Environment

In order to answer the research questions, we selected an elementary school site— Dr. Eric Jackman Institute of Child Study of the University of Toronto—where "Knowledge Building discourse" supported by Knowledge Forum technology (Scardamalia & Bereiter, 2003) is integral to the day-to-day work of the class. Students reference, evaluate, build on, and work to continually improve ideas—their own and those of community members. The quality of Knowledge Building discourse is increased through giving students greater collective responsibility for advancing ideas (Zhang, Scardamalia, Lamon, Messina and Reeve, 2007). Mirroring and extending their face-to-face discussions, students work in an online environment—Knowledge Forum (Scardamalia, 2002)—that provides a community space for collaborative work with ideas, the stored record of which constitutes data for the present research. Of particular relevance to this research are the following characteristics of Knowledge Forum: (a) "build-on," reference, linking, and annotation features that support interaction; (b) views that support graphical as well as textual literacy; (c) "rise-above" notes that facilitate the creation of higher-order knowledge structures; and (d) "scaffolds" (*i.e.* "My theory", "I need to understand", "New information") that make it easy for users and researchers to identify and tag contributor roles. These features enable research of a depth not possible with typical discourse environments, yet findings will be applicable to efforts to improve discourse in other on-line and face-to-face environments.

This study focuses on a Grade 4 inquiry of optics supported by Knowledge Forum. Both the teacher and students had multiple years of experience with Knowledge Building pedagogy and technology, so the situation represents what Fischer and Bidell (1997) call "optimal conditions" for identifying cognitive developmental goals. The study uses the same dataset as examined by Zhang et al. (2007), but focuses on individual ways of contributing to the collective discourse, to complement group-level variables. It explores students' patterns of contribution and examines how these patterns enhance (or impede) advancement of individual knowledge.

Participants and Dataset

The participants were 22 Grade 4 students (11 girls and 11 boys) and their teacher from the Dr. Eric Jackman Institute of Child Study, University of Toronto. The teacher has been committed to Knowledge Building and used Knowledge Forum for the previous two years. He worked collaboratively with students to identify problems of understanding and to discuss diverse ideas and theories through class face-to-face discourse. During the lessons, students were encouraged to conduct self-generated experiments, to search libraries as well as the Internet for interesting facts and to share new resources through cooperative reading. In parallel to these offline activities, students recorded their ideas, theories and findings in the Knowledge Forum database. This shared database helped students to keep their top-level questions in sight, while refining and improving newly generated theories. On a daily basis students were free to explore any problem related to the topic in question, and to contribute to the database.

The dataset analyzed below covers four months of online discourse on optics, including topics such as "How light travels," "Colors of light," "Natural and artificial light," "Shadows" and so forth (see Zhang et al., 2007 for details of the inquiry processes). A total of 318 notes have been analyzed in this study: 70% of all notes were students' personal notes (written by individual students), 20% were group notes (written by small groups of 2-5 students) and 6% were teacher's notes. This analysis examined only personal notes.

Analysis

Ways of Contributing List

On the basis of informal observation and the knowledge creation literature, a provisional list of contribution types was first created. This list was used to inform and guide our analysis of student online discourse that followed a grounded theory approach (Glaser & Strauss, 1967), with the categories elaborated, refined, then modified, and extended. Table 1 displays six major categories identified through the above process, consisting of 20 sub-categories in total.

Table 1: Ways of contributing to Knowledge Building dialogues.

Main category	Sub-category	Description of the category
Formulating thought-provoking questions	Formulating an explanatory question	Questions asking "why does it happen" and "how does it work"?
	Asking a design question	Questions asking "how can we prove/ test something?"

	Asking a factual question	Questions asking "what", "who" and "when"?
Theorizing	Proposing an explanation	Student proposes a theory that explains certain phenomena for the first time
	Supporting an explanation	Student supports an already existing theory (i.e. theory that has been proposed by another student) and provides a justification
	Improving an explanation	Student improves an already existing theory through elaborating, specifying details and using new evidence
	Seeking an alternative explanation	Student looks for a different explanation
Designing an experiment	Proposing/describing an experiment	Student proposes/describes an experiment to test an idea
	Identifying a design problem	Student describes an experiment that did not work, and identifies possible causes why it did not work
	Thinking of design improvements	Student tries to fix design problems and proposes a new/improved experiment
Working with evidence	Looking for evidence	Student asks or looks for evidence to support a particular idea
	Providing evidence or a reference to support an idea	Student provides evidence that comes either from his own experience or from authoritative sources to support a particular idea
	Providing evidence or reference to challenge or falsify an idea	Student provides evidence that comes either from his own experience or from authoritative sources to challenge or falsify a particular idea
	Finding new facts	Student brings up new interesting facts that do not support or reject a previous idea, but extend it
Creating syntheses and analogies	Synthesizing available ideas	Student synthesizes available ideas to create a better understanding of some phenomenon
	Creating analogies	Student moves to a higher level of understanding by creating an analogy to explain a particular phenomenon.
	Initiating a rise-above	Student summarizes previous ideas by integrating multiple notes into a rise-above note.
Supporting discussion	Using diagrams to communicate ideas	Student draws diagrams to communicate or support ideas.
	Giving an opinion	Student gives an opinion, but no explanation or justification
	Acting as a mediator	Student supports discussion by focusing on social roles rather than ideas.

Three independent raters used these categories to code the notes. When the same note fell into more than one contribution type, all related categories were counted. For example, if in the same note a student asked the question (e.g. "How light travels?"), proposed a theory to answer this question (e.g. "My theory is that light travels in a straight line") and included a diagram to make this theory visual, then three contribution types were counted: formulating an explanatory question, proposing an explanation, and using diagrams to communicate or support ideas. The three raters agreed in 80% of the coding. The remaining 20% were discussed to achieve consensus. Finally, in order to examine individual contributor patterns, the total number of contributions made by any individual student was calculated for each major category and sub-category.

Measures of Individual Knowledge Advancement

Individual knowledge advancement was assessed through content analysis (Chi, 1997) of student personal portfolio notes and a teacher-designed test. Each student wrote a portfolio note at the end of the optics inquiry to summarize what he/she had learned about light. Each portfolio note was divided into idea units, which were analyzed focusing on the scientific and epistemic levels of student ideas (see Zhang et al., 2007 for details).

Scientificness of student ideas was assessed using a four-point scale: 1–pre-scientific (containing a misconception while applying a naive conceptual framework), 2–hybrid (containing misconceptions that have incorporated scientific information), 3–basically scientific (containing ideas based on a scientific framework, but not precise) or 4–scientific (containing explanations that are consistent with scientific knowledge). This coding scheme was developed on the basis of Galili and Hazan's (2000) facets'-scheme framework for analyzing student optical knowledge.

Epistemic complexity of ideas was assessed using a rating scale with the following levels: 1–Unelaborated facts (containing simple description of terms, phenomena or experiences without elaboration); 2–Elaborated facts (containing elaborated description of terms, phenomena or experiences); 3–Unelaborated explanations (containing reasons, relationships or mechanisms without elaboration); 4–Elaborated explanations (containing elaborated versions of reasons, relationships, or mechanisms). To assess inter-rater reliability, two coders independently coded 12 portfolio notes: *Cohen's Kappa* = .83 for scientific sophistication and .75 for epistemic complexity (see Zhang & Messina, 2010 for details).

Teacher-designed test. The test was composed of 18 questions and administered at the end of the unit. It covered 10 of the 28 themes addressed in online discussions (e.g. shadows, rainbows.), with a full score of 54.

Results and Discussion

Distribution of Contributor Roles

Table 2 reports the frequency and percentage of each way of contributing identified from the student-generated optical discourse, including major and sub-category. Diverse ways of contributing were evident, connecting to and supporting one another for sustained, deepening discourse. The most frequent contribution types were those related to theorizing and working with evidence.

Table 2: Frequency of occurrence and percentages of contribution types (means and standard deviations).

Major category	N		%		Sub-category SD	N		%	
	M	*SD*	*M*	*SD*		*M*	*SD*	*M*	*SD*
Formulating questions	**3.73**	**2.66**	**17.89**	**8.14**	Explanatory questions	2.05	2.19	8.77	7.76
					Design questions	.14	.35	0.81	2.09
					Factual questions	1.55	1.14	8.31	7.11
Theorizing	**6.09**	**3.05**	**30.66**	**8.07**	Proposing an explanation	1.45	1.30	6.98	5.51
					Supporting an explanation	2.18	1.40	11.56	7.15
					Improving an explanation	1.55	1.74	7.25	6.63
					Seeking an alternative explanation	.91	1.15	4.86	5.95
Designing an experiment	1.05	1.00	5.27	5.06	Proposing/describing an experiment	.68	.78	3.42	4.00
					Identifying a design problem	.32	.57	1.76	3.15
					Thinking of design improvements	.05	.21	0.09	0.44
Working with evidence	**4.86**	**3.93**	**21.34**	**10.46**	Asking or looking for evidence	.23	.53	0.95	2.51
					Providing an evidence or reference to support a particular idea	1.55	1.41	7.13	5.33
					Providing an evidence or reference to challenge or falsify a particular idea	.55	.67	2.86	4.04
					Finding new facts	2.55	2.67	10.40	8.74
Creating syntheses and analogies	**1.18**	**1.30**	**5.74**	**6.19**	Synthesizing available ideas	.68	.78	3.38	3.60
					Creating analogies	.32	.57	1.35	2.55
					Initiating a rise-above entry	.18	.50	1.01	2.70
Supporting **discussion**	3.64	2.61	19.10	13.15	Using diagrams to communicate or support ideas	2.64	2.22	13.96	10.02
					Giving an opinion	.73	.88	3.93	5.59
					Acting as a mediator	.27	.55	1.21	2.49

Regression Analysis

Multiple linear regression analyses were performed to predict scientificness, epistemic complexity of ideas, and test scores from different contributor roles. Considering low occurrences observed for certain sub-categories, only major categories were used as candidate predictors, notably: questioning, theorizing, experimenting, working with evidence, creating syntheses and analogies, and supporting discussion.

Scientificness of Ideas

The overall regression, including 1 of the 5 candidate predictors, was statistically significant $R = .43$, $R^2 = .19$, adjusted $R^2 = .15$, $F(1, 20) = 4.57$, $p < .05$ (using a stepwise method). As indicated in Table 3, *working with evidence* accounted for about 19% of the variation in idea scientificness. The other five predictors did not enter the equation, failing to add to prediction based on scientificness alone.

Table 3: Results of statistical multiple regression (stepwise) to predict idea scientificness from six major contributor roles (N = 22).

	Zero-Order r		*b*	β	sr^2
	Scientificness	Working with Evidence			incremental
Working with evidence	.43*		.06*	.43	.186
Questioning	.19	.57**	—	—	—
Theorizing	.32	.77***	—	—	—
Experimenting	.19	.27	—	—	—
Creating syntheses and analogies	.13	.44*	—	—	—
Supporting discussion	.23	.19	—	—	—
			Intercept = 2.21***		
Mean	2.50	4.86			
SD	.54	3.93			

Analysis of results showed that *working with evidence* during Knowledge Building dialogue was a significant predictor of scientificness of individual ideas. More precisely, students who were actively searching authoritative sources, bringing new facts to discussions and using these facts to support or falsify ideas were able to increase scientificness of their own understanding. Thus, searching and processing the information for the community had a beneficial effect on individual knowledge.

Epistemic Complexity of Ideas

Multiple linear regression analysis was performed to predict epistemic level of ideas in portfolios from six major contributor roles. The overall regression, including 1 of the 5 candidate predictors, was statistically significant $R = .61$, $R^2 = .37$, adjusted $R^2 = .34$, $F(1, 20) = 11.76$, $p < .01$ (using stepwise method). As shown in Table 4, *theorizing* accounted for about 37% of the variation in the epistemic level of ideas. The other five predictors did not enter the equation, failing to add to prediction based on *theorizing* alone (even though most of these predictors had a significant zero-order correlation with it).

Table 4: Results of statistical (method = stepwise) multiple regression to predict epistemic complexity of ideas from six major contributor roles (N = 22).

	Zero-Order r		*b*	β	sr^2
	Epistemic Complexity	Theorizing			incremental
Theorizing	.61***		.05**	.61	.370
Questioning	.50**	.71***	—	—	—
Experimenting	.30*	.33	—	—	—
Working with evidence	.40*	.77***	—	—	—
Creating syntheses and analogies	.33	.39*	—	—	—
Supporting discussion	-.06	.10	—	—	—
			Intercept = 1.72***		
Mean	2.05	6.09			
SD	.27	3.05			

Analysis of results indicated several positive correlations between epistemic complexity of ideas in portfolios and major contributor roles played by participants during Knowledge Building discussions, i.e. theorizing, questioning, experimenting and working with evidence. However, only *theorizing* proved to be a significant predictor of epistemic complexity. More precisely, being dedicated to creating, supporting and improving explanations in a community context helped students to increase the epistemic complexity of their own ideas: portfolios of such "theorizers" contained elaborated reasons, relationships or mechanisms, instead of merely isolated facts.

Knowledge Test Scores

Multiple linear regression analysis was performed to predict knowledge test scores. The total *N* for this sample was 20; two cases were dropped due to missing data on at least one variable, and therefore for this analysis *N* = 20. The overall regression, including 1 of the 5 candidate predictors, was statistically significant $R = .51$, $R^2 = .26$, adjusted $R^2 = .22$, $F(1, 18) = 6.33$, $p < .05$ (using stepwise method). *Working with evidence* was the only significant predictor, explaining approximately 26% of the variance in the knowledge test scores (see Table 5 for details). Other five factors were not significantly predictive of the scores (even though there was a significant zero-order correlation for *theorizing*).

Table 5: Results of statistical (stepwise) multiple regression to predict knowledge test scores from six major contributor roles (N = 20).

	Zero-Order r		*b*	β	sr^2 incremental
	Post-Test Score	Working with Evidence			
Working with evidence	.51**		.49*	.51	.260
Questioning	.33	.53**	—	—	—
Theorizing	.39*	.72***	—	—	—
Experimenting	-.09	.19	—	—	—
Creating syntheses and analogies	.17	.31	—	—	—
Supporting discussion	-.12	.17	—	—	—
			Intercept = 40.96***		
Mean	43.25	4.86			
SD	3.65	3.93			

Correlation analysis indicated that there was a relationship between knowledge test scores and two major contributor roles: *theorizing* and *working with evidence.* Both roles were positively associated with content knowledge, but only the second one was significantly predictive of this knowledge. More particularly, active search of information in authoritative sources and exposure of this information to the community in order to support or falsify existing explanations helped students to increase their own content knowledge.

Conclusion

The goal of this study was to examine the extent to which contributing to a Knowledge Building online dialogue can predict individual advancement of knowledge in elementary school students. Toward this end, a systematic inventory of contributor roles was created including six major ways of contributing: questioning, theorizing, experimenting, working with evidence, creating syntheses and analogies, and supporting discussion. These roles were entered as candidate predictors of three individual knowledge measures: scientificness, epistemic complexity of ideas in portfolios and knowledge test scores. Analysis of results revealed two significant predictors: *working with evidence* and *theorizing*. The first one proved to significantly account for scientificness of ideas in student portfolios and knowledge test scores. Active use of empirical evidence and information from authoritative sources to support and develop ideas in community discourse led to high scores in both individual knowledge measures. The second predictor, *theorizing*, explained significant amount of variation in epistemic complexity of ideas in students' portfolios. More precisely, active generation, support and improvement of explanations during Knowledge Building discourse predicted high epistemic complexity of student understanding. Developing deeper explanations and examining them in light of information and knowledge from the larger field and collected evidence represent major ways of contributing to knowledge advancement in real world knowledge-creating communities (Bereiter, 2002; Sternberg, 2003). The present study confirmed the importance of fostering such contributions among students in order for them to work as productive Knowledge Building communities. Complimentary qualitative analysis of these ways of contributing is in progress. The combined findings will inform next steps of creating classroom interventions and technology designs (e.g., visualizing theory-evidence coherence, scaffolding, etc.) to support theorizing and working with evidence and authoritative sources for knowledge advancement. The sample size in this study was relatively small; future studies will use a larger sample of participants to re-examine the findings of this study, as well as identify new predictive relationships at the finer sub-category level.

The variety of ways of contributing to Knowledge Building dialogues that were examined in this study address knowledge creation and collaboration skills that lie at the core of 21st Century Skills (Partnership for 21st Century Skills, 2004). With all so-called "higher-order skills," a crucial and frequently ignored issue is their transferability (Bereiter & Scardamalia, 2006). Although it was beyond the scope of this study to examine transfer beyond the classroom, further research will be needed to examine transfer across different curriculum topics; positive results will lend support to the idea that developing distinctive ways of contributing can give students something transferable and helpful for them to become productive citizens in a knowledge society.

References

Bereiter, C., & Scardamalia, M. (2006). Education for the Knowledge Age: Design-centered models of teaching and instruction. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (pp. 695-711). Mahwah, NJ: Lawrence Erlbaum Associates.

Carnevale, A. P., & Desrochers, D. M. (2003). *Standards for what? The economic roots of K-16 reform* (Educational Testing Service Leadership 2003 Series). Princeton, NJ: Educational Testing Service. Available November 2008 at http://www.learndoearn.org/For-Educators/Standards-for-What.pdf

Chi, M. T. H. (1997). Quantifying qualitative analysis of verbal data: A practical guide. *Journal of the Learning Sciences, 6,* 271-315.

David, P. A., & Foray, D. (2003). Economic fundamentals of the knowledge society. *Policy Futures in Education*, *1*(1), 20-49. Retrieved July 18, 2004, from http://2_David_PFIE_1_1. Pdf

de Bono, E. (1985). *Six thinking hats*. Boston: Little, Brown.

Fischer, K. W., & Bidell, T. R. (1997). Dynamic development of psychological structures in action and thought. In R. M. Lerner (Ed.) & W. Damon (Series Ed.), *Handbook of child psychology: Vol 1. Theoretical models of human development* (5th ed., pp. 467–561). New York: Wiley.

Galili, I., & Hazan, A. (2000). Learners' knowledge in optics: Interpretation, structure and analysis. *International Journal of Science Education*, *22*(1), 57-88.

Glaser, B.G. & Strauss, A. (1967). The discovery of grounded theory: Strategies for qualitative research. New York: Aldine.

Hogan, K. (1999). Sociocognitive roles in science group discourse. *International Journal of Science Education, 21*(8), 855-882.

Homer-Dixon, T. (2006). The Upside of Down: Catastrophe, Creativity, and the Renewal of Civilization. Washington, DC: Island Press.

Partnership for 21st Century Skills (2004). *Collaboration skills*. Retrieved on September 20, 2008, from http://www.21stcenturyskills.org/index.php?option=com_content&task=view&id=263&Itemid=132

Resnick, M. (1996). Beyond the centralized mindset. *Journal of the Learning Sciences, 5*(1), 1-22.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Eds.), *Liberal education in a knowledge society* (pp. 76-98). Chicago: Open Court.

Scardamalia, M., & Bereiter, C. (2003). Knowledge building. In *Encyclopedia of education* (2nd ed. pp. 1370–1373). New York: Macmillan Reference.

Scardamalia, M., & Bereiter, C. (2007). Fostering communities of learners and knowledge building: An interrupted dialogue. In J. C. Campione, K. E. Metz, & A. S. Palincsar (Eds.), *Children's learning in the laboratory and in the classroom: Essays in honor of Ann Brown.* Mahwah, NJ: Erlbaum.

Sternberg, R. J. (2003). The development of creativity as a decision-making process. In R. K. Sawyer et al. (Eds.), *Creativity and development* (pp.91-138). New York: Oxford University Press.

Leng, J. Lai, M., & Law, N. (2008, October 28). *Characterizing patterns of interaction in knowledge building discourse.* Paper presented at the 16th International Conference on Computers in Education, Taipei, Taiwan.

Zhang, J., & Messina, R. (2010). Collaborative productivity as self-sustaining processes in a Grade 4 knowledge building community. In K. Gomez, J. Radinsky, & L. Lyons (Eds.), *Proceedings of the 9th International Conference of the Learning Sciences* (pp. 49-56). Chicago, IL: International Society of the Learning Sciences.

Zhang, J, Scardamalia, M, Lamon, M., Messina, R. & Reeve, R. (2007). Socio-cognitive dynamics of knowledge building in the work of 9- and 10-year-olds. *Educational Technology Research and Development*, *55*, 117-145.

Acknowledgments

This research was funded by a Research Grant from the Social Sciences and Humanities Research Council of Canada titled "Ways of contributing to dialogue in elementary school science and history." We would like to express our gratitude to the students, teachers and principals of the Dr. Eric Jackman Institute of Child Study, University of Toronto for the insights and the opportunities enabled by their involvement.

Knowledge Construction Patterns in Online Conversation: A Statistical Discourse Analysis of a Role-Based Discussion Forum

Alyssa Wise, Simon Fraser University, 250-13450 102 Avenue, Surrey BC, Canada, V3T 0A3 afw3@sfu.ca
Ming Ming Chiu, SUNY – Buffalo, 564 Baldy Hall, Buffalo, NY, mingchiu@buffalo.edu

Abstract: Assigning roles to individual students can influence the group's knowledge construction (KC) process during online discussions. Twenty-one students were divided into two groups and assigned rotating roles for eight one-week asynchronous online discussions. The KC contributions of all 252 posts in the discussion were coded using a five phase scheme; then statistical discourse analysis was applied to identify segments of discussion characterized by particular aspects of KC and "pivotal posts"—those posts which initiated new segments of discussion. Finally, the influences of assigned student roles on pivotal posts and KC were modeled. The results indicate that most online discussions had a single pivotal post separating the discussion into two distinct segments: the first dominated by a lower KC phase; the second dominated by a higher KC phase. The pivotal posts that initiated later segments were often contributed mid-discussion by students playing one of two summarizing roles (Synthesizer and Wrapper).

Introduction

Small group interactions in discussion forums are increasingly common during online learning and blended (face-to-face plus online) instruction. These discussions provide opportunities for negotiating meaning and co-constructing knowledge (Boulos & Wheeler, 2007). Like their face-to-face counterparts, online discussions can help learners participate in different aspects of the knowledge construction (KC) process: they can articulate their ideas and hear alternative points of view; debate and probe different ideas more deeply; synthesize and negotiate compromises in their positions; and test, modify and apply their newly-constructed knowledge (Stacey, 1999). However, simply putting students together in a computer-supported environment does not necessarily lead to knowledge construction (Fischer et al., 2002). Online conversations often remain exercises in listing ideas rather than engaging in rich interactions to construct shared understandings (Thomas, 2002). To support KC, online discussion environments must structure interactions to encourage productive collaboration (Dillenbourg, 2002).

We investigate one technique for supporting productive collaboration and KC during online discussions: assigning conversational roles (e.g., "Questioner") to script the process by which students interact (Hare, 1994). Specifically we use both content analysis and statistical discourse analysis (Chiu, 2008) to examine how: (a) the process of KC proceeds in an online discussion with assigned roles; and (b) a "summarizing" role in the middle of a discussion affects subsequent KC.

The Process of Knowledge Construction (KC)

The five-phase Interaction Analysis Model for Examining Social Construction of Knowledge developed by Gunawardena et al. (1997) is a tool for conceptualizing and assessing the process of KC during asynchronous discussions. The theoretical model was developed using a grounded theory approach and attempts to represent "the complete process of negotiation" involved in KC (p. 413). The model is presented in enough detail to guide empirical coding of the KC process and has been used extensively in studies involving roles.

In Gunawardena et al.'s (1997) model, KC occurs in a series of successive (though not necessarily strictly sequential) phases that can be viewed as moving generally from lower to higher mental functions. In the model, learners begin by sharing, clarifying, and elaborating ideas (Phase 1: *Sharing Information*). Then, conflicts among them are explored (Phase 2: *Exploring Dissonance*). Next, learners reconcile conflicting ideas by negotiating their meanings and co-constructing new knowledge propositions (Phase 3: *Negotiating Meaning*). Learners may then test and revise their synthesized ideas (Phase 4: *Testing and Modifying*). Finally, they can state and apply their new knowledge (Phase 5: *Agreeing and Applying*). The drive is to achieve higher phases of KC; however, successive phases build on each other. Hence, *all* phases contribute to the KC process (Gunawardena et al., 1997).

While Gunawardena et al.'s (1997) model conceptualizes knowledge construction as a process which occurs though learners' interactions (via their posts), previous work has not capitalized on its capacity to examine this process by analyzing *patterns* of KC. Past studies have often evaluated discussion quality by counting the posts in each KC phase (e.g., Schellens et al., 2007) or by computing each discussion's average KC phase (e.g., Schellens et al., 2005). These compiled measures treat KC as an outcome variable where more posts in higher KC phases indicate better learning, though the scheme was designed as a model of the *process* of constructing knowledge. Importantly, two discussions can proceed quite differently, while having the same KC phase counts (e.g., a sequence of posts with KC phases 1212312123 vs. 1111222233). By treating KC as an aggregate outcome of individual contributions, prior studies failed to test a central underlying premise of the model: groups construct knowledge through a specific sequence of phases. In this study we address this issue by analyzing how the group proceeds through the phases of the KC process.

Possible Patterns of KC Phases

Gunawardena et al.'s (1997) model suggests two possible KC patterns, but others are also theoretically possible. By identifying these patterns' characteristics, we can empirically test for them. These patterns are currently descriptive; future work will evaluate their benefits, drawbacks and relationship to KC outcomes.

Theoretically Predicted Pattern 1a: Strictly Progressive Segments for Each KC Phase

One interpretation of Gunawardena et al. (1997) views the KC phases as a strictly increasing sequence. Viewing KC as an interdependent process and a cumulative group effort, an individual's progress through the phases depends on and influences other group members, stimulating them to proceed through the phases more-or-less together. Transitions between the phases can thus be viewed as initiated by a "pivotal post," a contribution by a student (or the instructor) which changes the mode of discussion from one phase to another. This perspective aligns with other recent work on "pivotal moments" in the CSCL community (e.g. Lund et al., 2009).

A discussion that follows a strictly progressive sequence of the KC phases might proceed as follows. Initially, learners share ideas (a series of KC Phase 1 posts, e.g., 111). When a learner disagrees with another group member's idea (KC Phase 2), others may not always engage. Instead, they might continue proposing new ideas (e.g. KC phases 111211). In this case, the discussion continues in a sharing mode, identified by the dominance of posts in KC Phase 1 (with occasional posts in other phases). In contrast, a disagreement can act as a *pivotal post* that radically changes the mode of discussion. In this case because the pivotal post is a disagreement, the new mode of discussion becomes that of exploring disagreements (e.g. 11211→**2**22, the pivotal post is indicated in **bold**). At some point, a learner may attempt to reconcile views presented in different posts (KC Phase 3). This can provide a more cohesive view of disparate ideas—a common base around which group members can negotiate shared understandings. If others follow suit, the post serves as another pivotal post, and the group transitions from debating to reconciling ideas (e.g., 222→**3**3233). Next, a learner may start to test the negotiated idea(s) (KC Phase 4) which can stimulate more testing and revision of the idea(s), creating another higher phase discussion segment (e.g., 33233→**4**454344). Finally, if a learner formalizes and applies the revised idea(s), this can spark other applications in a KC Phase 5 discussion segment (e.g., 4454344→**5**5545). This hypothetical discussion follows Gunawardena et al.'s model of a "complete" knowledge construction process; it consists of five distinct, segments of progressively increasing KC phases with changes initiated by four pivotal posts (11211→**2**22→**3**3233→**4**454344→ **5**5545).

Theoretically Predicted Pattern 1b: Progressive and Regressive Segments for each KC Phase

Knowledge construction is not always a strictly linear process (Paavola et al., 2004). Thus another interpretation of Gunawardena et al. (1997) recognizes earlier phases as logically prior to later phases, but also allows regressive segments: segments dominated by lower KC phases than the previous segment. For example, a tentatively shared synthesis (KC Phase 3) might break down when a learner returns to debating the merits of a particular idea (KC Phase 2) and others follow suit (e.g., 11211→**2**22 →**3**3233→**2**212). In this pattern, any number of segments can occur, and the return to a "lower" phase as part of the KC process is not necessarily negative for the discussion.

Alternative Patterns

There are other theoretical alternatives to the ones Gunawardena et al. (1997) suggest. One possibility is that groups might skip one or more KC phases. For example, learners might share their ideas (e.g., 111211) and then propose a compromise among them (111211→**3**333) without exploring differences or disagreements (skipping KC Phase 2). Then, they might conclude their discussion without testing or applying it (Phases 4 and 5). Here, the discussion has only two segments (111211→**3**333), though each segment is dominated primarily by posts in one KC phase. This pattern of segments with skipped KC phases can be strictly progressive (Pattern 2a) or include regressive segments (Pattern 2b; e.g. 111211→**3**333→**2**2122). In both cases the patterns do not require passing through all earlier phases to reach later ones. Other alternatives are that the KC phases of posts increases through progressive segments that are not clearly defined by a single phase (Pattern 3: e.g. 111→**2**21312132) or that no distinct segments of KC phases exists (Pattern 4: e.g. 15314215). Next, we discuss how assigned student roles and the functions they ask learners to perform align with the KC phases and might influence the above KC patterns.

Supporting KC in Online Discussions with Assigned Student Roles

Online learning conversations often do not realize their potential as sites of rich KC. One way to increase the likelihood of valuable learning interactions is by assigning roles to students to script their collaboration (Dillenbourg & Hong, 2008). Roles give students guidance about how to interact with one another productively in ways that promote desired cognitive, metacognitive and socio-cognitive processes (King 2007). Common assigned roles include moderator, starter, wrapper, planner, theoretician, source searcher, responder and summarizer (e.g., Schellens et al., 2005; Strijbos et al., 2004).

Roles can support collaborative KC in online discussions by creating positive interdependence and mutual accountability among students (Schellens et al., 2007; Strijbos et al., 2004), leading to increased interaction (Hara et al., 2000; Seo, 2007) and integration of discourse (Persell, 2004). Roles can also support students' metacognitive awareness of their contributions to the group's KC (Persell, 2004; Strijbos et al., 2004), helping them to self-moderate discussions and increase their autonomy, ownership, motivation and responsibility for learning (Seo, 2007). However, not all roles influence posts' KC, and specific assigned roles can have different influences (for example, contrast the positive effects of the "Wrapper" role with the negative effects of the "Source Searcher" role in De Wever et al., 2007 and Schellens et al., 2005; 2007).

While past research suggests that some particular roles can have a positive impact on KC during online discussions, research gaps remain. In particular, researchers have not examined the interdependent nature of the group processes underlying KC, in which each learner's posts help build the context for others' future posts. Specifically, work is needed to investigate how role-based posts influence other participant's postings and overall group KC patterns. To consider how specific roles may interact with the KC process, we can examine the alignment between the KC phases and the specific functions that each role asks a learner to enact. If a role asks a learner to perform a function that aligns with a KC phase different from the KC phase of the group members' current posts, that contribution could act as a pivotal post that initiates discussion in a new KC phase. Building on De Wever et al.'s (2007) efforts to assess role enactment, Wise et al., 2009) analyzed assigned roles in the literature and identified six core conversational functions that they ask learners to perform: Introduce New Idea, Bring in Source, Use Theory, Respond, Give Direction, and Summarize. We propose that some of these functions conceptually align with specific KC phases. Specifically, we focus our attention on the Summarize function and its associated roles because it theoretically aligns with advanced phases of knowledge construction (De Wever et al., 2007; Schellens et al., 2005; 2007).

Learners in Summarizing Roles May Create Pivotal Posts That Advance KC

Summarize is a synthetic function that asks a learner to organize and integrate different ideas in the discussion (Wise et al., 2009). For example, in a discussion on math lesson planning, a learner can describe how several different suggested activities for teaching parallel lines could be combined. Cognitively this helps the summarizing learner (and potentially those who read the post) to consolidate their understanding of different ideas. From a socio-cognitive perspective, summarizing posts can support the group in building on the existing discussion, maintaining joint attention and coordinating their activity.

With respect to the KC process, a summarizing post can identify areas of agreement and aid compromises between contested ideas (which aligns theoretically with KC Phase 3, Negotiation of Shared Meaning). Or, it can help group members reach a final agreement and recognize changes in their ideas (which aligns with KC Phase 5, Statement/ Application). While empirical studies show that summarizing *posts* consistently contribute at a high KC phase (De Wever et al., 2007; Schellens et al., 2005; 2007), *groups* assigned roles with a summarizing function have not consistently outperformed those without one (e.g., compare Schellens et al., 2005 and 2007). One reason other group members may not realize the benefits of summarizing posts is that the Summarize function is often assigned to a Wrapper role asked to conclude a discussion. (Schellens et al., 2005; 2007; see also Hara, et al., 2000). Since a Wrapper generally summarizes at the end of a discussion, other group members are unlikely to make subsequent posts and thus realize the coordination benefits described above.

We propose using the Summarize function in the middle of a discussion to synthesize discussion strands, maintain joint attention and ground subsequent discussion. While few students are likely to post after the Wrapper at the end of a discussion, many more are likely to build on a midway summary that helps them integrate their understanding of the various ideas. Reading a summarizing post can prompt them to join the synthetic effort to co-construct shared understandings with their groupmates. As more students do this, they create more posts in higher KC phases, compounding these processes until the whole group collectively climbs to a higher phase of KC. In this way, a midway summarizing post is potentially a pivotal post that can elevate a group to a higher KC phase, helping to solve the insidious problem of online discussions stuck in the rut of simply listing ideas without developing them collaboratively (Thomas, 2002).

Research Questions

Past work has looked at KC in aggregate, losing important information about patterns of KC and how they may be influenced by the assignment of student roles. In addition, the beneficial Summarize function has been primarily assigned to roles that post summaries at the end of discussions, limiting its potential to elevate group KC processes. In this study, we used a temporal analysis to examine KC as a process in a series of role-based discussions. One particular role ("Synthesizer") was used to elicit a Summary midway through each discussion.

We asked the following questions:

1. What pattern(s) characterize KC processes during an online, asynchronous discussion with assigned roles?
2. Does a summary midway through the discussion affect subsequent KC?

Methods

Participants

Twenty-one students (8 women, 13 men) taking a blended (face-to-face and online) Educational Technology course. Seven of the 21 participants were of Asian descent. There were 10 undergraduate students, 8 graduate students, and 3 practicing teachers. Class members formed two discussion groups: (a) undergraduates and (b) graduate students and teachers.

Learning Context & Procedure

For each of the eight weeks of online asynchronous discussion, groups were asked to collectively create a solution to an authentic instructional design challenge (e.g. in one week they were asked to design a set of activities to help 10 year-olds become "experts" in the Chinese Zodiac). Each week, students were given one of 10 functionally-aligned roles to play: Starter (New Idea, Give Direction), Inventor (New Idea), Importer (New Idea, Bring in Source), Mini-me (Use Theory), Elaborator (Respond), Questioner (Respond), Devil's Advocate (Respond), Traffic Director (Given Direction), Synthesizer (Summarize, Give Direction), Wrapper (Summarize). These were randomly assigned and rotated so that each student played a different role in each of the eight weeks. Roles were introduced to students via a role description guide and modeled by the instructor in an earlier practice discussion. Students were required to post twice (first five weeks) or once (final three weeks) each week in the LMS Moodle discussion tool (http://moodle.org/); discussion participation was 15% of the course grade. A total of 252 posts were created and collected for content analysis.

Content Analysis

Function enactment and KC were evaluated at the level of the individual post by two independent coders for all posts in the data set. Krippendorff's α (Krippendorff, 2004) was used to assess inter-rater reliability.

Enactment of Functions

The Analysis Scheme Identifying Message Characteristics-Functional (Wise et al., 2010) has six dimensions corresponding to the six conversational functions. Krippendorff's α was: New Idea (.65), Bring in Source (.92), Use Theory (.73), Respond (.98), Give Direction (.76), and Summarize (.88). New Idea was coded as Absent or Present; other dimensions were coded as Absent, Minorly Present, or Extensively Present.

Knowledge Construction (KC)

Gunawardena et al.'s (1997) scheme was used to code the highest KC phase achieved in each post (α= .84).

Statistical Analysis

The KC in each group's weekly online discussions was modeled at the micro- and meso-levels using Statistical Discourse Analysis (SDA) (Chiu, 2008). As described below, SDA first statistically determined pivotal posts ("breakpoints" in statistical terminology) and discussion segments based on the KC phase exhibited in posts, then tested explanatory models for these pivotal posts, and finally tested explanatory models for the KC phase of an individual post. Several levels of variables were used in the explanatory models to capture the characteristics of student demographics, learning activities, roles and posts. The statistical power of any regression (including SDA) for this sample size of 252 posts is .99 for an effect size of .30 (α = .05; Cohen et al., 2003).

In the initial phase of analysis, for each week of each group, we statistically identified pivotal posts that divided discussions into segments by using regression analysis to model different possible numbers and locations of pivotal posts and finding the model that best fit the data (Chiu & Khoo, 2005). First, we modeled the KC phase of each post in a discussion under the assumption of no pivotal posts. Next, we assumed one pivotal post and tested all possible locations of the pivotal post. Then, we assumed two pivotal posts and tested all possible pairs of locations of two pivotal posts, and so on up to all possible location combinations for six pivotal posts. The best model of pivotal posts had the smallest Bayesian information criterion (BIC, also known as Schwarz information criteria, Kennedy, 2004).

In the second phase of analysis we created an explanatory model (a multilevel, binary logit regression) to identify characteristics associated with pivotal posts and the posts preceding them. We entered variables at multiple levels (e.g., activities, student characteristics, and post characteristics) in sets according to time constraints, expected causal relationships, and likely association with pivotal posts. If the variable Summarize was significant, we tested whether the effects differed across the type of summarizing by replacing the Summarize variable with Minor Summary and Extensive Summary. We then used lag variables to model the characteristics of the preceding posts. We first entered a set of post characteristics for the previous post (lag 1), then added the same explanatory variables for the post before that (lag 2). No variables were significant at lags greater than 2. To test for moderation effects by discussion segment, we added terms for the interaction of the above variables with a variable indicating posts after the first pivotal post. A nested hypothesis test (χ^2 log likelihood test, Kennedy, 2004) with an alpha level of .05 was used to test whether each set of explanatory variables was significant. Multi-level mediation tests were used to test if a variable M mediated an X→Y relationship: X→M→Y (Krull & MacKinnon, 2001).

Finally, in the third phase of analysis we created an explanatory model for the KC phase of an individual post. As above, we entered factors at multiple levels (e.g., activities, student characteristics, and post characteristics). The procedure was the same as that of modeling pivotal posts with the following exceptions. First, KC has 5 ordered values (1, 2, 3, 4, 5), so multilevel, ordered logit was used rather than multilevel, binary logit. Second, to examine differences in KC across discussion segments we added a variable that indicated posts created After the 1st Pivotal Post. Third, in the Current Post Characteristics variable set, we did not include KC since it is the outcome variable.

Results

Descriptive Statistics

Results confirmed that discussions were enacted without overrepresentations of posts from specific roles, from specific students or in particular weeks. The ten assigned roles yielded equal numbers of posts, except for the Starter and Wrapper roles which should and did make only one corresponding functional post per discussion. Over half the posts were in KC Phase 1 (Sharing Information), with substantial numbers in KC Phase 3 (Negotiating Meaning) and KC Phase 5 (Agreeing and Applying). Few focused on discrepancies or contradictions (KC Phase 2 [Exploring Dissonance] or KC Phase 4 [Testing and Modifying]).

Identifying Pivotal Posts

The discussions averaged one pivotal post each (two time segments) and most segments had a majority of posts in a single KC phase (see Figure 1). The Synthesizer and Wrapper roles contributed most of the pivotal posts, and these pivotal posts often were Extensive Summaries in KC phases 3 or 5. Other pivotal posts had varied characteristics.

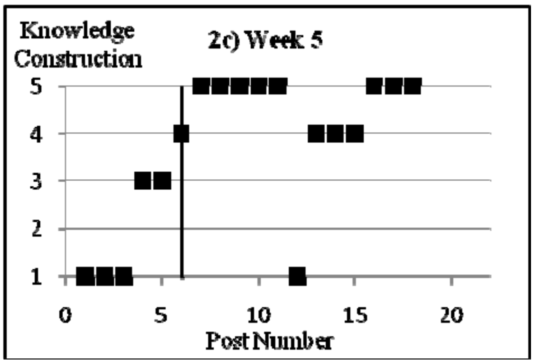


Figure 1. Example of a Discussions with 1 Pivotal Post (Yielding 2 Segments of Discussion).

Modeling Pivotal Posts

Assigned role and current post characteristics accounted for a substantial portion of the pivotal post variance (28%, see Figure 2). Compared to other roles, Synthesizers' and Wrappers' posts were more likely to be Extensive Summaries, and Extensive Summaries were more likely than other functions to be pivotal posts. Other variables (including discussion group and post order in the discussion) were not significant. Notably, Minor Summaries were not significantly more likely to be pivotal posts.

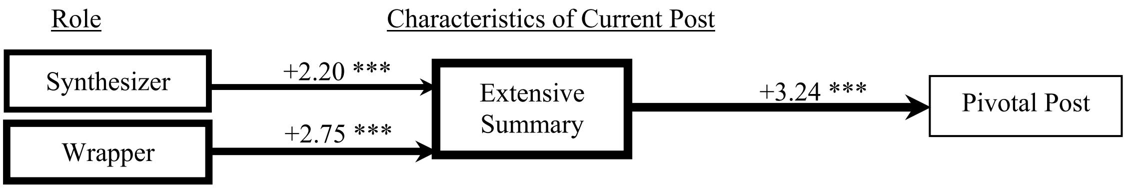


Figure 2. Path diagram of final model predicting pivotal posts. Numbers shown are regression coefficients. Solid lines indicate positive effects. Thicker lines indicate larger effect sizes. *p<.05, **p<.01, ***p<.001.

Modeling the KC Phase of Posts

Discussion segment, role, characteristics of the current post, and characteristics of the two prior posts accounted for much of the variance in the KC phase of a post (see Figure 3). Only 23% of the differences in KC phases occurred across discussions; 77% of the differences were within each discussion. Across all discussions, Synthesizers and Wrappers posted many more Summaries compared to other roles. Posts with Summaries (Minor or Extensive) exhibited a higher KC phase on average. Thus, Synthesizer and Wrapper posts averaged higher KC phases compared to other roles with the effect largely mediated by the Summary function.

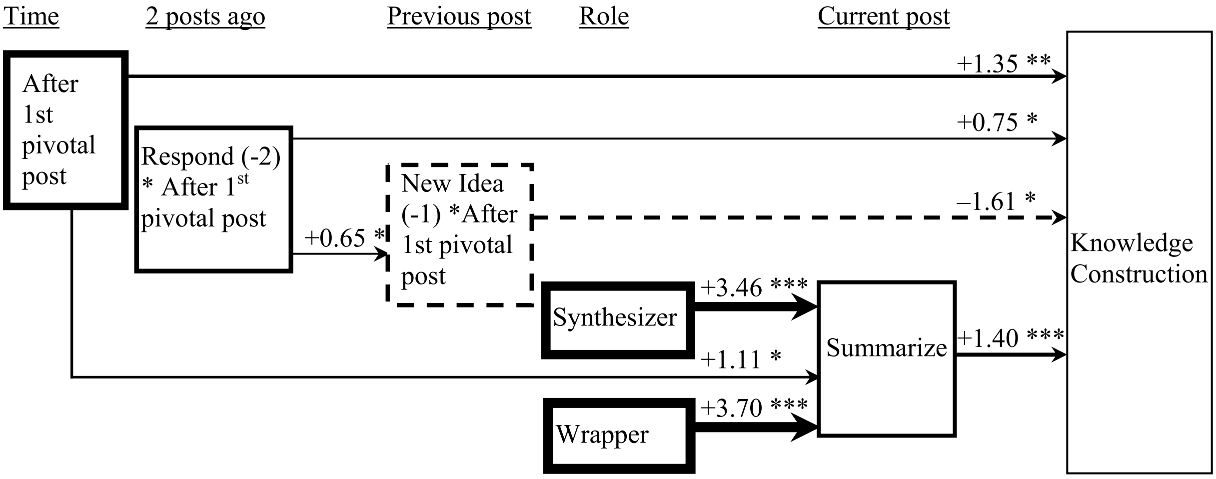


Figure 3. Path diagram of final model predicting knowledge construction. Numbers shown are regression coefficients. Solid lines indicate positive effects. Dashed lines indicate negative effects. Thicker lines indicate larger effect sizes.*p<.05, **p<.01, ***p<.001.

The results also show three time-specific relationships. First, the KC phase was substantially higher in posts after a pivotal post than before one. Second, Summaries were more likely to occur after a pivotal post. Third, there were two links between explanatory variables and KC phase that existed only after a pivotal post has occurred: one, if the previous post had a New Idea, the current post averaged a lower KC phase; and two, if a post was Responsive, the following post was more likely to be a New Idea and the KC phase was slightly lower two posts later (indirect and direct effects combined). All other variables were not significant. Notably, the order of a post in the discussion did not affect the KC phase of a post, showing that later posts did not necessarily show more advanced phases of KC. Furthermore, neither discussion group nor week explained a significant amount of variance in the KC phase of post.

Discussion

The study revealed a distinct KC pattern that emphasized sharing (Phase 1), negotiating (Phase 3) and summarizing (Phase 5) ideas, but not dissonance (Phase 2) or testing tentative syntheses (Phase 4). A pivotal post divided most discussions into two distinct segments. Segments were generally characterized by a majority of posts in a single KC phase with later segments showing higher KC than earlier segments. The pivotal posts that initiated later segments were typically contributed by students assigned the Synthesizer or Wrapper roles and contained Extensive Summaries that elevated KC both immediately and in subsequent posts. Below we discuss these results with respect to our research questions and the previous literature.

Research Question 1: What Pattern(s) Characterize the KC Process During An Online Asynchronous Discussion with Assigned Roles?

Like previous studies of online discussions with assigned roles, our analysis allowed us to examine the proportions of posts in each KC phase. As in prior work (Gunawardena et al., 1997; Schellens et al., 2005), most posts in this study were in KC Phase 1 (Sharing Information) showing that students produced new ideas much more often than they considered existing ideas. In addition, similar to some prior findings (Schellens et al. 2007), this study also showed a greater proportion of posts in KC Phase 3 (Negotiating Meaning) than in Phase 2 (Exploring Dissonance). This is different from a pattern in which the proportion of posts decreases for each successive KC phase (e.g. De Wever et al., 2007). In comparison with past studies, the learners in this study had many more posts in KC Phase 5 (Agreeing and Applying), though still few posts in KC Phase 4 (Testing and Modifying). The preponderance of posts in the initial (1) and convergent (3 and 5) phases and lack of posts in judgmental KC phases (2 and 4) suggest that these groups were focused on reaching consensus.

Moving beyond aggregate counts of posts, our analysis probed the process of knowledge construction by identifying segments of KC patterns. This let us empirically test the underlying premise that groups construct knowledge through a specific sequence of phases such as Gunawardena et al.'s (1997) theoretically proposed patterns of KC. Most of the online discussions studied had at least two distinct segments of discussion (one pivotal post), which rejects the KC Pattern 4 hypothesis (no distinct segments of KC). Segments were generally characterized by a majority of posts in a single KC phase, with later segments showing higher KC; this rejects the KC Patterns 1b and 2b hypotheses (which include regressive segments). In particular, the discussions often had an initial segment with mostly KC Phase 1 posts (Sharing Information) followed by a statistically identified pivotal post which elevated the discussion to a segment with a majority of posts in KC phases 3 or 5. This rejects KC Pattern 1a (which requires a segment for each KC phase) and provides strong support for the KC Pattern 2a hypothesis (strictly progressive segments with some KC phases skipped). However, nearly a quarter of the later segments did not have a majority KC phase which also provides some support for the KC Pattern 3 hypothesis (mixed KC phase segments). These results differ from the two theoretically "complete" patterns suggested by Gunawardena et al. (1997; KC Patterns 1a and 1b), both which include segments for all five phases of KC. Notably, in the original conceptualization, exploration of dissonance and testing a proposed synthesis are important to KC, but these results show that in some cases, groups can engage in KC Phase 5 processes without KC Phases 2 or 4 in this context. Importantly, while some disagreeing posts in KC Phases 2 and 4 were made, these contributions did not propel the group into a critique-focused segment of discussion. The lack of disagreements in the discussions may be due to concerns about social relationships (Chiu, 2008b), inadequate concern for the quality of the solution, or a notion of agreement as an indicator of a quality solution.

Whether the absence of disagreements affects the quality of knowledge construction outcomes is an important question both empirically and theoretically and at both group and individual levels. At the group level, exploration of dissonance between ideas is thought to be important as a foundation for quality co-construction of knowledge. At the individual level, the cognitive dissonance caused by engaging with conflicting ideas is needed to trigger a learner to reconsider existing ideas and construct new understandings. In future work, we will explore these issues by empirically studying how characteristics of the discussion process (such as a lack of posts in Phases 2 and 4) influences independent learning outcome measures at both the group and individual level.

Research Question 2: Does a Summary Midway through the Discussion Affect Subsequent KC?

Like past studies, our results show that roles encouraging summarization (Synthesizer, Wrapper) yielded posts in significantly higher KC phases than posts by roles which did not (De Wever et al., 2007; Schellens et al., 2005; 2007;). More importantly, our temporal analysis allowed us to examine the effects of these summarizing posts on the group's patterns of collaborative KC. As hypothesized, mid-discussion extensive summaries created by students in the Synthesizer role were often pivotal posts that initiated new discussion segments with posts in elevated phases of KC. Due to a large number of late student posts, many Wrapper summaries inadvertently ended up mid-discussion and also acted as pivotal posts that advanced the KC phase of the discussion. Notably, only one post with a minor summary was a pivotal post, suggesting that minor summaries are qualitatively different from extensive summaries.

These results suggest that reading the extensive summaries facilitated learner contributions at higher KC phases on average, thus advancing the group's KC process. Particularly, the integrative value of extensive summaries (Hara et al., 2000) can help students consolidate their understanding of the different ideas contributed and draw on the previous discussion to negotiate shared understandings (KC Phase 3) or apply their newly-constructed knowledge (KC Phase 5). In this way, the extensive summary can coordinate group activity and ground subsequent discussion. This result illustrates the power of our temporal analysis to illuminate how individuals' posts in a discussion can influence group processes of knowledge construction.

Conclusion

This study showed differences in KC phases of posts across time in an asynchronous online discussion. Notably, most discussions had only two discussion phases, unlike Gunawardena, et al.'s (1997) 5-phase Interaction Analysis Model. While past studies have shown that students benefit from writing a summary and that this can be encouraged using assigned roles, this study used a temporal analysis to show that extensive summaries mid-discussion can also change the tenor of the group's conversation and elevate it to higher KC phases. Future work will examine how different KC discussion patterns can affect group and individual learning outcomes.

References

Boulos, M. N., & Wheeler. S. (2007). The emerging web 2.0 social software: An enabling suite of sociable technologies in health and health care education. *Health Information and Libraries Journal, 24*, 2–23.

Chiu, M. M. (2008). Flowing toward correct contributions during groups' mathematics problem solving: A statistical discourse analysis. *Journal of the Learning Sciences, 17* (3), 415 - 463.
Chiu, M. M., & Khoo, L. (2005). A new method for analyzing sequential processes: Dynamic multi-level analysis. *Small Group Research, 36,* 600-631.
Cohen, J., West, S. G., Aiken, L., & Cohen, P. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Lawrence Erlbaum.
Dillenbourg, P. & Hong, F. (2008). The mechanics of CSCL macro scripts. *International Journal of Computer-Supported Collaborative Learning, 3*(1), 5-23.
Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. In P. A. Kirschner (Ed.) *Three worlds of CSCL: Can we support CSCL?* (pp. 61–91). Heerlen: Open Universiteit Nederland.
De Wever, B., Van Keer, H., Schellens, T., & Valcke, M. (2007). Applying multilevel modeling to content analysis data. *Learning and Instruction, 17(*4), 436-447.
Fischer, F., Bruhn, J., Gräsel, C., & Mandl, H. (2002). Fostering collaborative knowledge construction with visualization tools. *Learning and Instruction*, *12*(2), 213-232.
Goldstein, H. (1995). *Multilevel statistical models*. Sydney: Edward Arnold.
Gunawardena, C. N., Lowe, C. A., & Anderson, T. (1997). Analysis of a global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing. *Journal of Educational Computing Research, 17*(4), 397-431.
Hara, N., Bonk, C. J., & Angeli, C. (2000). Content analyses of on-line discussion in an applied educational psychology course. *Instructional Science, 28*(2), 115-152.
Hare, A. P. (1994). Types of roles in small groups: A bit of history and a current perspective. *Small Group Research, 25*(3), 433-448.
Kennedy, P. (2004). *A guide to econometrics*. Cambridge: Blackwell.
King, A. (2007). Scripting collaborative learning processes: A cognitive perspective. In F. Fischer, I. Kollar, H. Mandl, & J. Haake (Eds.), *Scripting computer supported communication of knowledge: Cognitive, computational and educational perspectives* (pp. 13-37). New York: Springer.
Krippendorff, K. (2004). *Content analysis*. (2nd ed.).. Thousand Oaks, CA: Sage.
Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research*, *36,* 249-277.
Lally, V. (2001). Analysing teaching and learning interactions in a networked collaborative learning environment: Issues and work in progress. In *Euro CSCL 2001* (pp. 397–405). Maastricht McLuhan Institute.
Lund, K., Law, N., Rosé, C., Suthers, D., & Teplovs, C. (2009). *Pinpointing pivotal moments in collaboration*. Workshop held at the STELLARnet Alpine Rendez-Vous, Garmisch-Partenkirchen Germany.
Paavola, S., Lipponen, L., & Hakkarainen, K. (2004). Models of innovative knowledge communities and three metaphors of learning. *Review of Educational Research, 74,* 557-576.
Persell, C. H. (2004). Using focused web-based discussion to enhance student engagement and deep understanding. *Teaching Sociology, 32*, 61-78.
Schellens, T., Van Keer, H., & Valcke, M. (2005). The impact of role assignment on knowledge construction in asynchronous discussion groups: A multilevel analysis. *Small Group Research*, *36*(6), 704-745.
Schellens, T., Van Keer, H., De Wever, B., & Valcke, M. (2007). Scripting by assigning roles. *International Journal of Computer-Supported Collaborative Learning, 2* (2/3), 225-246.
Seo, K. K. (2007). Utilizing peer moderating in online discussions: Addressing the controversy between teacher moderation and nonmoderation. *American Journal of Distance Education, 21(*1), 21-36.
Sheskin, D. J. (1997) *Parametric and nonparametric statistical procedures*. Boca Raton, FL: CRC Press.
Stacey, E. (1999). Collaborative learning in an online environment. *Journal of Distance Education*, *14*(2),14-33.
Strijbos, J. W., Martens, R. L., Jochems, W. M. G., & Broers, N. J. (2004). The effect of functional roles on group efficiency. *Small Group Research, 35*, 195-229.
Thomas, M. J. W. (2002). Learning within incoherent structures: The space of online discussion forums. *Journal of Computer Assisted Learning, 18*, 351-366.
Wise, A. F., Saghafian, M. & Padmanabhan, P. (2009). Comparing the functions of different assigned student roles in online conversations. *Proceedings of the E-Learn Conference 2009—World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education,* Vancouver, BC.
Wise, A. F., Saghafian, M. & Padmanabhan, P. (2010, May). *ASIMeC-F: A content analysis scheme for assessing the presence of conversational functions in asynchronous discussions*. Paper presented at the Annual Meeting of the American Education Research Association, Denver, CO.

Quantified Measures of Online Discourse as Knowledge Building Indicators

Jianwei Zhang, Yanqing Sun, University at Albany
Email: jzhang1@albany.edu, ysun@albany.edu

Abstract: This secondary data analysis examined a set of social interaction (e.g., social network patterns), content (e.g., questions, ideas), and lexical measures (e.g., academic words, domain terms) applied to a Knowledge Forum discourse database created by 22 fourth-graders as they investigated optics over a four-month period. Knowledge advancement was evaluated based on student portfolio notes focusing on the depth and breadth of their optical understanding. Correlations found between the measures of social interaction, content, and lexical usage in the discourse and the depth and breadth of student understanding help to empirically justify a set of online discourse measures that are sensitive to knowledge productivity. The results suggest a framework to inform the selection, creation, and integrated use of online discourse measures in research as well as design of automated assessment tools embedded in collaborative learning environments.

Introduction

The field of computer-supported collaborative learning (CSCL) faces the challenge to develop methodologically justifiable measures of collaborative knowledge building as a distributed and emergent process driven by students' diverse input. Intensive effort has been made to analyze and assess student online discourse using quantitative and qualitative methods (e.g., de Laat, Lally, Lipponen, & Simons, 2007; Guzdial & Turns, 2000; Koschmann, 2001; Meier, Spada, & Rummel, 2007; Stahl, 2006; Suthers, Dwyer, Median & Vatrapu, 2010). Quantified measures are diverse, analyzing participation rate, social network patterns, vocabulary use, content contributions, and so forth. Choices of research measures for a given study are often made based on theoretical considerations of what such measures mean and imply. There is a need to examine and justify the importance of these measures to collaborative knowledge building through systematic empirical testing, which will further provide a stronger research base for initiatives to create automated analysis tools (e.g., Rosé et al., 2008). To address this need, the present study applied a range of quantified measures to the same discourse database and examined their relationship to student knowledge advancement.

Knowledge building—the creation of knowledge as a social product through collective and sustained efforts—becomes pervasive in a knowledge-based society (Bereiter, 2002). Recent educational initiatives thus emphasize engaging students in collaborative knowledge building with the support of new technology environments. Achieving this goal requires educational changes from individual to collaborative processes and outcomes; from teacher-designed to student-driven goals and processes; from a focus on content coverage to that on depth of understanding; and from standard learning outcomes to student diverse expertise (Brown, Ash, Rutherford, Nakagawa, Gordon, & Campione, 1993; Scardamalia, 2002; Stahl, 2006; Zhang, Scardamalia, Lamon, Messina, & Reeve, 2007). Traditional assessment tools focus on individual learning processes and outcomes based on predefined learning objectives and curriculum standards. Although these tools can also be used in knowledge building and CSCL research; they reveal very little, if any, about collaborative processes, emergent and progressive understanding, and community knowledge advancement (Zhang & Chan, 2008). Thus, CSCL researchers face the challenge to develop and integrate new research measures and assessments.

Various research measures have been developed in the CSCL literature to analyze and assess collaborative knowledge building, often using student discourse as a primary data source. Three types of quantified measures have been widely used: (a) Content analysis (Chi, 1997), using coding schemes to categorize the nature of responses, types of questions, depth of ideas, evidence use, argumentation patterns, and so forth (e.g., Baker et al., 2007; Hakkarainen, 2003; Hmelo-Silver, 2003; van Aalst & Chan, 2007; Weinberger & Fischer, 2006; Zhang et al., 2007); (b) Socio interaction measures, focusing on contribution rate, reading rate, conversation threads (build-on trees), social networks of who reads or responds to whose postings (Aviv, Erlich, Ravid, & Geva, 2003; de Laat, Lally, Lipponen, & Simons, 2007; Guzdial & Turns, 2000; Hewitt & Teplovs, 1999; Hewitt, Brett, & Peters, 2007; Howell-Richardson & Mellar, 1996; Zhang, Scardamalia, Reeve, & Richard, 2009); and (c) Linguistic markers of discourse, such as occurrences of epistemic words and domain-specific key terms in discourse (Hong & Scardamalia, 2008; Sun, Zhang, & Scardamalia, 2010). These measures have been informed by various CSCL theories and models to capture important aspects of collaborative learning and knowledge building. However, the significance of these measures in relation to student knowledge advancement has rarely been systematically examined, partly because different researchers tend to use different measures to analyze CSCL discourse from their own point of view (Hmelo-Silver, 2003).

On the basis of online discourse measures, efforts are made to create computer-based research tools to automate some of the analyses, such as using text classification technology to automate or aid content analysis (Law et al., 2007; Rosé et al., 2008), analyzing patterns of participation and interaction based on user log files (e.g., Burtis, 1998), and extracting, comparing, and clustering key terms used in online discourse through semantic analysis (Teplovs & Fujita, 2009). Automated assessment tools are further designed and embedded in collaborative online environments to provide teachers and their students with concurrent feedback as their work proceeds (Scardamalia, Bransford, Kozma, & Quellmalz, 2010). With data mining and other computing technologies easing data analysis, the challenge becomes what to analyze and how to combine the different measures and rich amount of data based on a sound framework so researchers, teachers, and students can make meaningful interpretations and informed decisions.

In our recent research, we developed and adapted a set of research tools to examine collaborative knowledge building supported by an online environment. These included inquiry thread analysis for mapping out communal knowledge growth by identifying and tracing discourse contributions to different problem spaces (Zhang et al., 2007), social network analysis for evaluating collaboration and collective responsibility (Zhang et al., 2009), content analysis of student contributions and depth of understanding (Zhang et al., 2007, 2009), and lexical analysis of student discourse to examine the growth of productive written vocabulary in relation to scientific understanding (Sun et al., 2010). These measures were applied to the same dataset—an online discourse database of a Grade 4 classroom focused on optics. The goal of the present study was to conduct a secondary analysis of the above measures to identify significant indicators of knowledge building, which can inform the selection of CSCL research measures and the design of automated analysis and assessment tools.

Method

The Knowledge Building Context

The participants were 22 fourth-graders (11 girls and 11 boys) from an elementary school in downtown Toronto. This study analyzed their inquiry of optics conducted over a four-month period in line with principles and practices of knowledge building, supported by Knowledge Forum (Scardamalia & Bereiter, 2006). Knowledge Forum provides a communal, multimedia knowledge space, represented as different *views* (workspaces) corresponding to students' knowledge building goals. Students contribute *notes* to views to share and continually advance their ideas, using a set of interaction tools (e.g., build-on, rise-above, referencing) to engage in knowledge building discourse (see Figure 1 for a screenshot). Both the students and their teacher had multiple years of experience with knowledge building pedagogy and Knowledge Forum.

Figure 1. Knowledge Building Discourse in the Colors of Light View. Each small square icon represents a note, and a line connecting two notes represents a build-on.

During the optical inquiry, the fourth-graders generated problems of understanding, discussed diverse ideas and theories through face-to-face knowledge building discourse, conducted self-generated experiments and observations, searched libraries and the Internet, and shared new resources through cooperative reading. Along with these offline activities, they shared their questions, ideas, data, and information sources in Knowledge Forum for sustained discourse that extended and enriched their classroom conversations. The teacher experimented with having the whole class collaborate opportunistically to understand optics and to progressively identify important, related issues (e.g., light sources, how light travels, colors, lenses and mirrors, vision) to deepen the inquiry. Knowledge Forum provided the public space in which their collective works were recorded, with new views created in line with emergent goals and linked to existing views. These interconnected views helped to keep the top-level goal center front and to keep the structure fluid: sub-goals were identified and elaborated in related views and small groups formed and reformed based on evolving needs. On a daily basis, students were free to explore any problem from any view in the database. They all took responsibility for the overall growth of the database. Near the end of the inquiry, each student wrote a reflective portfolio note to summarize what he/she had learned about light. Analyses of student portfolios and a pre- and post-test demonstrated productive advancement of knowledge (see Zhang et al., 2007 for details).

Measures of Online Discourse

The primary data source was student discourse in Knowledge Forum. Over four months, students created 287 notes in seven views (e.g., Shadows, Colors of Light, How Light Travels). The optical discourse database was a proportion of the data analyzed in several related studies (Zhang et al., 2007, 2009; Sun et al., 2010). This secondary analysis focused on three sets of measures that had been applied to this database, including social interaction measures, content-based coding, and lexical analyses. These measures are summarized in Table 1 and elaborated below.

Table 1: Measures of online knowledge building discourse.

Category	Measures	Explanations
Social interaction measures	Note contribution	# of notes authored per student, as an indicator of their contrition to the community space.
	Note reading percentage	% of notes read, as an indicator of knowledge sharing and information spread.
	Note reading network: in-degree and out-degree	Social network of who reads whose notes, with in-degree and out-degree indicating the extent to which a member receives and sends out note-reading contacts from/to different members, respectively.
	Note linking network: in-degree and out-degree	Social network of who links to whose notes through build-on, rise-above, and reference citation, with in-degree and out-degree indicating the extent to which a member receives and sends out note-linking contacts from/to different members, respectively.
	Note linking network: cliques	In a social network of who links to whose notes, a clique is a sub-network of members who have more note linking ties to each other than to members who are not part of the group. The number of cliques each student belongs to indicates the level of dynamic collaboration and idea contact.
Content measures	Problems	# of notes raising and addressing deepening problems about the topic.
	Personal ideas	# of notes that contributed student understanding and claims.
	Information sources	# of notes rephrasing or summarizing information from readings, the Internet, the teacher, parents, etc.
	Evidence	# of notes that test and justify ideas using experiments, observations, or life experiences.
	Inquiry threads	An inquiry thread is a conceptual stream of discourse that addresses a shared principal problem. The number of inquiry threads each student contributes to as an author indicates diverse participation in the community's knowledge space.
Lexical measures	Total words	Total words written per student in the online discourse.
	1st 1,000 words	Percentage of the 1st 1000 most frequent English word families used in student notes, as an indicator of limited vocabulary and writing.
	Academic words	The percentage of academic words (e.g., theory, hypothesis, approach) used in student notes, as an indicator of productive academic discourse.
	Domain-specific words	Student use of domain-specific words as an indicator of their appropriation of disciplinary discourse and knowledge.

Among other analyses of social interactions (e.g., note contribution, reading), we adopted a set of measures from social network analysis (Wasserman & Faust, 1994). In a social network, each community member is represented as a node, and a relational tie (e.g., build-on) between two members as a line. We used social network analysis to examine two types of social relations recorded by Knowledge Forum: (a) who read whose notes, with reading peers' notes as a primary mean to understanding knowledge advances and challenges of the community; and (b) who linked to whose notes (i.e., created build-ons, rise-aboves, or references), as a indicator of complementary and connected contributions. Three measures were included in this analysis: (a) in-degree showing how many relational ties (e.g., reading, linking) a member received from peers, suggesting the level of his/her influence; (b) out-degree measuring the number of relational ties one sent out to other peers as an indicator of his/her effort to understand and build on peer contributions; and (c) clique analysis, which identified sub-networks each member belonged to in the note linking network, as an indicator of community-wide dynamic collaboration. A clique is "a sub-set of a network in which the actors are more closely and intensively tied to one another than they are to other members of the network." (Hanneman, 2001, p. 79)

Content analysis (Chi, 1997) was adopted to code: (a) questions identified by students in their notes (e.g., how do solar panels work?); (b) student personal ideas that presented their own theories and claims often labeled as "My theory" (e.g., "If there is no light, there can't be a shadow"); (c) information sources, to introduce new information from readings, the Internet, the teacher, or parents, etc., often labeled as "New information," and use the information to deepen their understanding; (d) evidence, to examine and deepen their understanding using findings from experiments and observations; and (e) inquiry threads contributed to. An inquiry thread consists of a series of discussion entries that address a shared principal problem and constitute a conceptual line of inquiry in a community knowledge space (Zhang et al., 2007). These entries may involve multiple physical threads of build-ons. For example, students in this study wrote 27 notes in an extended discussion about how rainbows are created, constituting an inquiry thread titled "Rainbows," with deeper questions progressively addressed leading to improved understanding. Within the communal knowledge space, 28 inquiry threads were identified, each beginning with the first note created and ending with the last note created or modified (see Zhang et al., 2007 a visual representation). Students engaged in the inquiry themes through opportunistic interactions based on their interest. Tracing student notes contributed to different inquiry threads helped to examine their emergent, diverse participation in the community's knowledge space.

Increasing use of sophisticated, low frequency words in free writing indicates growth of productive vocabulary and writing skills (Nation 2001). Thus, lexical frequency analysis was employed to examine student use of three types of words in their online discourse: (a) The first 1,000 most frequent word families in English (West, 1953). Low-proficiency writers tend to rely more on these basic word families in writing; (b) A list of academic words, including 570 word families that are typical of academic discourse across disciplinary areas, enabling references to other authors and findings (e.g., *assume, establish, conclude*) and processing of data and ideas (e.g., *analyze, assess, category*) (Coxhead 1998). Writers need to gain productive written control of the academic vocabulary in order to be recognized as a member of the academic discourse community (Corson, 1997); and (c) Domain-specific terms, which included 89 domain words related to light (e.g., names of optical concepts, devices and phenomena) identified from the Ontario Curriculum (Sun et al., 2010).

Assessing Knowledge Gains Based on Student Portfolio Notes

Assessing student understanding based on reflective essays or portfolios has been tested in a number of studies (e.g., van Aalst & Chan, 2007; Zhang et al., 2007). This study analyzed student portfolio notes to assess their knowledge gains focusing on two aspects: knowledge diffusion and depth of understanding.

Knowledge diffusion (or idea spread) becomes an important issue in learning contexts that encourage diverse participation and distributed expertise (Brown et al., 1993). Our analysis thus examined whether individual students could benefit from the community's knowledge advances in diverse inquiry themes to achieve adequate breadth of understanding beyond their personal focus. Specifically, the first author segmented each student's portfolio note into idea units—the smallest unit of text that conveyed a distinct idea about light. Each idea unit was coded in relation to the inquiry threads (themes) that emerged from the knowledge building discourse (e.g., how light travels, nature of shadows, eclipses, rainbows) (see Zhang et al., 2009 for details).

To look at the depth of student understanding, each idea unit was additionally rated in terms of epistemic complexity and scientific sophistication. Epistemic complexity indicates students' efforts to produce not only descriptions of the material world, but also theoretical explanations and articulation of hidden mechanisms, which are central to the focus of science (Salmon, 1984). A four-point scale (1 - unelaborated facts, 2 – elaborated facts, 3 – unelaborated explanations, and 4 - elaborated explanations), adapted from Hakkarainen's (2003) work, was used to code each idea unit (for details and examples, see Zhang et al., 2007). Scientific sophistication focuses on the extent to which a student has moved from an intuitive toward a scientific framework. It is gauged through coding students' ideas in their portfolio notes based on a four-point scale (1 - pre-scientific, 2 - hybrid, 3 - basically scientific, and 4 - scientific) (for details, see Zhang et al., 2007), which was informed by Galili and Hazan's (2000) facets-scheme framework for analyzing students' misconceptions in optics. To assess inter-rater reliability, two coders independently coded 12 portfolio notes, *Cohen's Kappa* = .83 for scientific sophistication, *Cohen's Kappa* = .75 for epistemic complexity (Zhang & Messina, 2010).

Epistemic complexity represents the level of complexity at which a student chooses to approach an issue. Scientific sophistication represents the level of success a student has achieved in processing an idea at a certain complexity level. It is easier to convey a scientific idea at a factual level (e.g., "there are different colors in a rainbow"), but harder to provide a scientific explanation of a fact (e.g., elaborate what causes a rainbow and why the colors are always in the same order). The meaning of the scientific score of an idea is dependent on the level of its complexity. Therefore, a composite score was used to indicate the depth of understanding by multiplying the above two ratings, weighting the rating of scientific sophistication with the level of complexity (Zhang et al., 2009). For example, an idea rated as "1 - unelaborated facts" and "3 – basically scientific" will have a composite score of 3, while an idea rated as "4 - elaborated explanations" and "3 – basically scientific" will have a composite score of 12.

Results

To identify quantified measures of online discourse that may have a strong connection with student knowledge productivity, we calculated the correlations between these measures and the depth and breadth of student understanding, which represent two independent components of the learning outcome with virtually no correlation ($r = -.03$).

Social Interaction Measures

Table 2 reports the correlations between social interaction measures of the online discourse and the depth and breadth of student optical understanding gauged based on their portfolio notes. Student deep understanding of optics was associated with high rates of note contribution and note reading—both reading others' notes and being read by others—in the knowledge building discourse, with significant ($p < .05$) or marginally significant correlations ($p < .10$). Two of the social network measures of note linking contacts are significantly correlated with the depth of understanding achieved ($p < .05$). Students with deeper understanding received more intensive note linking contacts from their peers and collaborated with multiple sub-networks of students through building on, rising above, and referencing one another's work. There is a close to significant correlation between students' in-degree in the note linking networks and the breadth of understanding achieved ($p < .10$), showing that students who understood a broader range of issues had received more note-linking contacts from peers in the knowledge building discourse.

Table 2: Correlations (Pearson r and p) between social interaction measures of online discourse and student optical understanding.

	Notes written	% of notes read	Note reading network: in-degree	Note reading network: out-degree	Note linking network: in-degree	Note linking network: out-degree	Note linking network: Cliques belonging to
Depth of understanding	.437* (.042)	.398† (.067)	.519* (.013)	.398† (.067)	.431* (.045)	.214 (.338)	.469* (.028)
Breath of understanding	.198 (.377)	.105 (.644)	.308 (.164)	.061 (.788)	.364† (.096)	-.068 (.765)	.159 (.478)

Note. † $p < .10$, * $p < .05$

Content-Based Measures

As Table 3 shows, student deep understanding is significantly ($p < .05$) or marginally significantly ($p < .10$) correlated to their efforts to generate and contribute personal ideas, identify and address deeper problems, and incorporate informative sources to help them better understand light. Not surprisingly, the breadth of their understanding achieved is strongly correlated to the number of inquiry threads—each addressing a principal problem—they contributed to during the optical discourse.

Table 3: Correlations (Pearson r and p) between the content-based measures of online discourse and student optical understanding.

	# of notes identifying problems	# of notes contributing personal ideas	# of notes incorporating new sources	# of notes using evidence	# of inquiry threads/themes contributed to
Depth of understanding	.582** (.004)	.365† (.095)	.403† (.063)	.260 (.242)	-.034 (.879)
Breadth of understanding	.296 (.182)	.288 (.193)	-.009 (.970)	.056 (.806)	1.000*** (.000)

Note. † $p < .10$, ** $p<.01$, *** $p<.001$

Lexical Measures

Table 4 displays the correlations between the lexical measures of online discourse and the depth and breadth of student optical understanding. The depth of their understanding is positively correlated to the total number of words students wrote and the occurrences of domain-specific words and academic words in their online notes. Student knowledge productivity is associated with their engagement in online written discourse that incorporates a larger number of domain-specific words in optics (e.g., *shadow, reflect, absorb, wave*) and epistemic, academic words that are characteristic of academic discourse (e.g., *hypothesis, conclusion*). There is a significant negative correlation between the depth of student optical understanding and the occurrence of the most basic, 1^{st} 1,000 English word families in the online discourse, which indicates a limited level of vocabulary and writing. A significant positive correlation was found between the breadth of understanding and the number of unique domain words students used in their notes. Spontaneous incorporation of domain-specific words in online discourse suggests the expanding scope and richness of inquiry in a domain.

Table 4: Correlations (Pearson r and p) between the lexical measures of online discourse and student optical understanding.

	Total words written	% of the 1^{st} 1,000 words	% of the academic words	# of unique domain words	Total domain words
Depth of understanding	.646** (.001)	-.646** (.001)	.506* (.016)	.458* (.032)	.660** (.001)
Breadth of understanding	.250 (.262)	-.302 (.172)	.226 (.313)	.594** (.004)	.218 (.329)

Note. * $p<.05$, ** $p<.01$

Discussion

This study investigated a set of social interaction, content-based, and lexical measures applied to the same knowledge building discourse database. Examining their correlations with the depth and breadth of student understanding helped to identify and justify indicators of online discourse conducive to knowledge building. Several social interaction measures indicate productive discourse to achieve deep understanding, including the number of notes contributed, percentage of notes read, in-degree (being read by peers) and out-degree (reading peers' work) in the note reading network, and in-degree (being built on by peers) and dynamic memberships in cliques (sub-networks) in the note linking network developed through build-ons, rise-aboves, and referencing citations of peer ideas. Content-based discourse indicators associated with student deep understanding involve the number of notes contributing personal theories, identifying deepening problems, and incorporating new information sources, with student contributions to multiple inquiry threads strongly connected to the scope of their optical understanding. The number of notes reporting evidence is not significantly correlated to the depth of student understanding, possibly because this analysis only considered the frequency of evidence use. Additional measures might examine how evidence was used to support reasoning and discourse. Finally, all the lexical discourse measures have significant correlations to the depth of student understanding, including total words written, occurrences of academic words and domain-specific words (both total and unique words), and less frequent use of the 1st 1,000 English word families. Incorporating unique domain-specific words in the knowledge building discourse additionally suggests the expanding scope and breadth of inquiry.

The above-identified measures collectively characterize productive knowledge building discourse along four interrelated dimensions. (a) *Interactive engagement in extended discourse*, with community members understanding and successively building on to one another's intellectual input over time beyond short-threaded conversation turns (Engle, 2006; Guzdial & Turns, 2000; Suthers et al., 2010; Zhang et al., 2007, 2009). Students are thus expected to have a high note contribution rate, note reading percentage, in-degree and out-degree in the note reading network, and in-degree and clique memberships in the note linking network. (b) *Idea-centered, progressive discourse*, with students engaging in idea generation and improvement, expanding a shared base of knowledge, and identifying deeper challenges as their understanding deepens, harnessing sophisticated language tools to communicate and develop ideas (Bereiter, 2002; Hakkarainen, 2003; Hmelo-Silver, 2003; Zhang et al., 2007; Zhang & Messina, 2010). Such efforts are evident when students actively identify deepening questions, propose initial theories for peer input, and develop better theories and explanations. (c) *Constructive use of knowledge and language resources*, through making constructive use of authoritative sources and appropriating academic vocabulary and discourse (e.g., academic words, domain concepts) in the related domain areas to support idea development (Chernobilsky, DaCosta, & Hmelo-Silver, 2004; Hong & Scardamalia, 2008; Sun et al., 2010; Zhang et al., 2007). (d) *Student-driven, dynamic* *collaboration*, with students identifying progressive goals and forming dynamic teams to address challenges emerged at the intersections of their interests. Analyses in this regard may examine student-generated questions and goals, distributed network patterns, and emergence of cliques in note link networks (Zhang et al., 2009). Efforts along the above dimensions help foster collective responsibility for knowledge advancement in a community (Scardamalia, 2002; Zhang et al., 2009).

In conclusion, this study provides empirical justification for a set of quantified measures used in the CSCL literature. The three types of measures that capture four dimensions of knowledge building discourse may be used as a framework to guide the selection and integrated use of research measures in specific contexts and development of new analyses to capture the interactional, cognitive, and linguistic processes of knowledge building. This framework can be further elaborated and used to guide the design of automated assessment and feedback tools in collaborative learning environments. Focusing on these and additional indicators, automated assessment tools may provide coherent and accessible analyses of the interactional, cognitive, and linguistic processes to aid student reflection on and improvement of knowledge building. Automated analyses focusing on various aspects of online discourse can be integrated to address pedagogically valuable questions, such as: Is there interactive engagement in extended discourse? To what extend are we improving our ideas and what contributions are evident? Are we engaging in productive writing and discourse? Are we enacting collective responsibility for knowledge building?

This investigation of online discourse measures was based on a small sample of 22 students, which has prevented us from conducting confirmative factor analysis to further test how the different indicators capture different dimensions of productive online discourse. Although this study focused on quantified measures only, complementary qualitative analyses have been reported elsewhere that helped to elaborate and contextualize the quantified patterns through detailed accounts (Zhang et al., 2007; Zhang & Messina, 2010). Further analysis will include additional measures of student knowledge growth (e.g. post-test scores) and examine correlations across the three types of knowledge building indicators.

References

Aviv, R., Erlich, Z., Ravid, G., & Geva, A. (2003). Network analysis of knowledge construction in asynchronous learning networks. *Journal of Asynchronous Learning Networks, 7*(3) (an online journal).

Bereiter, C. (2002). *Education and mind in the knowledge age*. Mahwah, NJ: Erlbaum.

Brown, A.L., Ash, D., Rutherford, M., Nakagawa, K., Gordon, A., & Compione, J. C. (1993). Distributed expertise in the classroom. In G. Salomon (Ed.), *Distributed cognitions* (pp.188-228). New York, NY: Cambridge University Press.

Burtis, J. (1998). *Analytic Toolkit report*. Toronto, ON: Ontario Institute for Studies in Education of the University of Toronto.

Chernobilsky, E., DaCosta, M. C., & Hmelo-Silver, C. E. (2004). Learning to talk the educational psychology talk through a problem-based course. *Instructional Science, 32*(4), 319-356.

Chi, M. T. H. (1997). Quantifying qualitative analysis of verbal data: A practical guide. *Journal of the Learning Sciences, 6*, 271-315.

Corson, D.J. (1997). The learning and use of academic English words. *Language Learning, 47*, 671-718.

Coxhead, A. (1998). An academic word list. *Occasional publication Number 18, LALS*, Victoria University of Wellington, New Zealand.

de Laat, M., Lally, V., Lipponen, L., & Simons, R.J. (2007). Investigating patterns of interaction in networked learning and computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning, 2*, 87-103.

Engle, R. A. (2006). Framing interactions to foster generative learning: A situative explanation of transfer in a community of learners classroom. *Journal of the Learning Sciences, 15*(4), 451-498.

Galili, I., & Hazan, A. (2000). Learners' knowledge in optics: Interpretation, structure and analysis. *International Journal of Science Education, 22*(1), 57–88.

Guzdial, M. & Turns, J. (2000). Computer-supported collaborative learning in engineering: The challenge of scaling up assessment. In M. J. Jacobson & R. B. Kozma (Eds.), *Innovations in science and mathematics education: Advanced designs for technologies of learning* (pp. 227-257). Mahwah: NJ: Lawrence Erlbaum Associates.

Hakkarainen, K. (2003). Progressive inquiry in a computer-supported biology class. *Journal of Research in Science Teaching, 40*(10), 1072-1088.

Hanneman, R. A. (2001). *Introduction to social network methods.* Riverside, CA: Department of Sociology, University of California, Riverside.

Hewitt, J., & Teplovs, C. (1999). An analysis of growth of patterns in computer conferencing threads. In C. Hoadley (Ed.), *Proceedings of the third international conference on computer support for collaborative learning* (pp. 232–241). Mahwah, NJ: Erlbaum.

Hewitt, J., Brett, C. & Peters, V. (2007). Scan Rate: A new metric for the analysis of reading behaviors in asynchronous computer conferencing environments. *American Journal of Distance Education, 21*(4), 1-17.

Hmelo-Silver, C.E. (2003). Analyzing collaborative knowledge construction: Multiple methods for integrated understanding. *Computers & Education, 41*, 397-420.

Hong, H.-Y., & Scardamalia, M. (2008). Using key terms to assess community knowledge. *Paper presented at the Annual Meeting of American Educational Research Association*, New York, NY.

Howell-Richardson, C., & Mellar, H. (1996). A methodology for the analysis of patterns of participation within computer mediated communication courses. *Instructional Science, 24*, 47–69.

Koschmann, T. (2001). Revisiting the paradigms of instructional technology. *Proceedings of the 18th Annual Conference of the Australian Society for Computers in Learning in Tertiary Education*. Melbourne.

Law, N., J. Yuen, Huang, R., Li, Y. & Pan, N. (2007). A Learnable Content and Participation Analysis Toolkit for Assessing CSCL Learning Outcomes and Processes. *Paper presented at the International Conference on Computer Supported Collaborative Learning*, July 16-21, New Jersey.

Meier, A., Spada, H., Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration process. *International Journal of Computer-Supported Collaborative Learning, 2*, 63-86.

Nation, I.S.P. (2001). *Learning vocabulary in another language.* Cambridge, UK: Cambridge University Press.

Rosé, C., Wang, Y.-C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., Fischer, F. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in CSCL. *International Journal of Computer-Supported Collaborative Learning, 3*(3), 237-271.

Salmon, W. C. (1984). *Scientific explanations and the causal structure of the world*. Princeton, NJ: Princeton University Press.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Eds.), *Liberal education in a knowledge society* (pp. 67-98). Chicago, IL: Open Court.

Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In K. Sawyer (Eds.), *Cambridge Handbook of the Learning Sciences* (pp. 97-118). New York, NY: Cambridge University Press.

Scardamalia, M., Bransford, J., Kozma, R., & Quellmalz, E. (2010). New assessments and environments for knowledge building. *Assessment and Learning of 21st Century Skills*. Paper posted to http://www.atc21s.org/home/

Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge*. Cambridge, MA: MIT Press.

Sun, Y., Zhang, J., & Scardamalia, M. (2010). Knowledge building and vocabulary growth over two years, Grades 3 and 4. *Instructional Science, 38*(2), 247-271.

Suthers, D., Dwyer, N., Medina, R. & Vatrapu, R. (2007). A framework for conceptualizing, representing, and analyzing distributed interaction. *International Journal of Computer-Supported Collaborative Learning, 5*(1), 5-42.

Teplovs, C., & Fujita, N. (2009). Determining curricular coverage of student contributions to an online discourse environment through the use of latent semantic analysis and term clouds. *Proceedings of the 9th international conference on Computer supported collaborative learning (Vol. 2*, pp. 165-167). International Society of the Learning Sciences.

van Aalst, J., & Chan, C.K.K. (2007). Student-directed assessment of knowledge building using electronic portfolios. *The Journal of the Learning Sciences, 16*(2), 175-220.

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education, 46*, 71-95.

West, M. (1953). *A general service list of English words*. London, UK: Longman, Green & Co.

Zhang, J., & Chan, C.K.K. (2008). Examining the growth of community knowledge in an online space. In *Workshop Proceedings of the International Conference on Computers in Education (ICCE 2008)* (pp. 60-69). The Asia-Pacific Society for Computers in Education.

Zhang, J., & Messina, R. (2010). Collaborative productivity as self-sustaining processes in a Grade 4 knowledge building community. In K. Gomez, J. Radinsky, & L. Lyons (Eds.), *Proceedings of the 9th International Conference of the Learning Sciences* (pp. 49-56). Chicago, IL: International Society of the Learning Sciences.

Zhang, J., Scardamalia, M., Lamon, M., Messina, R., & Reeve, R. (2007). Socio-cognitive dynamics of knowledge building in the work of nine- and ten-year-olds. *Educational Technology Research and Development, 55*(2), 117–145.

Zhang, J., Scardamalia, M., Reeve, R., & Messina, R. (2009). Designs for collective cognitive responsibility in knowledge building communities. *Journal of the Learning Sciences, 18*(1), 7–44.

Critical Moments of Knowledge Co-construction: Reconsidering Meaning-making of Postings in Online Group Discussion

Hsiu-Mei Chang, Graduate Institute of Learning & Instruction, National Central University, auroraspp@gmail.com
Fei-Ching Chen, Graduate Institute of Learning & Instruction, National Central University, fcc@cc.ncu.edu.tw
Kai-Tien Cheng, Graduate Institute of Atmospheric Sciences, National Central University, joseph0220@me.com
Ren-Yow Tzeng, Graduate Institute of Atmospheric Sciences, National Central University, ren@atm.ncu.edu.tw

Abstract: This study attempts to investigate the critical moments in online knowledge co-construction. We analyzed the episodes of the dynamic group discussion by reconsidering the meaning-making of each posting in order to gain a full contextual understanding of how collective knowledge developed. The results identified three critical moments: "confusion and hesitation", "from stick around to move forward" and "making justification and breaking through". Unlike using the single posting or adjacent pairs of postings as the unit of analysis, this study claims that linking backwards to earlier turns is as important as linking forwards to later turns for discovering a complete picture of how online group members co-construct knowledge. The nuance of dealing with the time dimension across different approaches in CSCL is also discussed.

Introduction

Learning through knowledge co-construction is increasingly regarded as having a significant effect (e.g., Stahl, 2009). Ideally, through observing and responding to each other, group members establish a focus for their discussion and maintain a heterogeneous environment of revision and complement within which knowledge is deepened and augmented. Practically, however, it is not easy to achieve a positive effect. In the process of group collaboration, members encounter a lot of problems, such as: having different levels of domain knowledge, becoming distracted from on-topic discussion, struggling in coordination with other members, logging on sporadically etc. The quality of the knowledge co-construction in group discussion is therefore decreased.

To investigate this online knowledge building phenomenon, many researchers used the "single posting" as the unit of analysis (Suthers, Dwyer, Medina & Vatrapu, 2007) to more easily scrutinize visible messages in order to identify the quality of knowledge co-construction in group discussion. Others treated the "each thread" as the unit of analysis in order to capture the response relationship in the group discussion, the formation of the sub-thread, the lifespan of a thread, the frequency and type of certain behavior of the group members, and so on (e.g., Hewitt, 2005). This study extends this line of research by exploring exactly how the "whole thread" is analyzed and what aspects of knowledge co-construction can be discovered.

In this study, we analyze the discussions in the Learning Atmospheric Sciences via InterNet, LAIN, an independent and voluntary discussion forum. Unlike the related literature, we purposely use both the "single posting" and "the whole thread" as units of analysis in order to compare and contrast the distinct differences found in capturing the relationships between the members, and the roles of the group members, respectively. With these, we then attempt to address the following issues: First, to identify the critical moments of knowledge co-construction in heterogeneous online group discussion, and second, to determine how the heterogeneous group members collaboratively accomplish knowledge co-construction.

Analytical Approaches to Knowledge Co-construction

"Knowledge co-construction" is a promising issue addressed by different parties with similar terms in the CSCL field. For example, knowledge building (Scardamalia, 2002), knowledge convergence and collaborative learning (Jeong & Chi, 2007) and group cognition (Stahl, 2010). Knowledge building communities have to tackle the difficult problem of handling complexity and reaching a common understanding. By the same token, approaches to knowledge building groups or communities are also diverse and multi-voiced. Two issues have been the focus: 1) proposing elements in capturing the phenomenon of knowledge co-construction, and 2) proposing analytical approaches in analyzing interaction among knowledge co-constructors.

First, rather than strive to understand the process by which knowledge co-construction developed, most previous studies on knowledge co-construction attempted to investigate the elements of knowledge co-construction. In particular, these studies explored those that contributed to the success or failure of group collaboration and identified certain elements: question and response construction, engagement in the group process (Zhou, 2009), knowledge sharing and the willingness to revise (Chiu, 2008), the focus of discussion, the connection between old and new ideas, the association of the discussion with the topic (e.g., Aalst, 2009), knowledge convergence and knowledge divergence (Jeong & Chi, 2007), the heterogeneous group member (e.g., Scardamlia & Beretier, 2006), the peer as each member's reader or writer (e.g., Dennen & Wieland, 2007), the application of information, and references or authoritative materials (Scardamalia & Briereter, 2006; Hmelo-Silver, 2006; Stahl, 2008). Such lists are helpful in grasping partially the knowledge co-construction in group discussion. However, we argue that an authentic process of knowledge co-construction in heterogeneous group discussion could be represented not just as isolated incidents in interaction episodes, but also as a mixture of coherent and incoherent processes. It is unlikely to be possible to separate knowledge co-construction from the process of dynamic interaction when all elements in this context are so thoroughly intertwined. In the present study, we therefore attempt to broaden our view in order to capture the critical moments and the contextually situated processes of group knowledge co-construction.

Second, many researchers have devoted a great deal of effort to analyzing the quality of knowledge co-construction. We therefore trace the development of the analytic method during the last decade: In the early period, previous research adopted the following analytic methods to measure learning in discussion forums: (a) restricting subject behaviors; (b) setting certain variables; (c) coding and counting the content of "single posting", "adjacency pairs" and "taking turns"; and (d) the quantitative-statistics. Lately, however, some researchers claim that the above methods hinder appreciation of sequential structure and ignore the nature of collective group growth (e.g., Suthers, Dwyer, Medina & Vatrapu, 2007). Besides which, the order, sequentiality and timing typically play a significant role in how the postings are understood (Stahl, 2010). Therefore, from the perspective of sequential and situated interaction, researchers have developed particular concepts and methods, such as statement analysis, interaction analysis, episode analysis (e.g., Jeong & Chi, 2007; Zhou, 2009; Reimann, 2009), uptake analysis (Suthers, 2006), CORDTRA Diagrams (Hmelo-Silver, Chernobilsky, & Mastov, 2006), inter-subjectivity and group cognition (e.g., Stahl, 2010) and prior posting (Diggelen, Janssen, & Overdijk, 2008). Some have used the multi-threads or cross threads as the unit of analysis in order to understand the formation, structure, growth, and lifespan of the thread, and the synergetic interaction and the interaction pattern between the threads (e.g., Schrire, 2006; Hewitt, 2005). Along with this line of research approach, our analytic method focuses on the sequence of contextual events, and on the meaning-making of individual postings in the process of knowledge co-construction in heterogeneous online group discussion.

Specifically, in an attempt to highlight the distinct difference in the effect of capturing the nature of knowledge co-construction between the "contextual-situated process" and the "single posting", we purposely apply two different units of analysis to the same dataset. First, using the "single posting" as the unit of analysis, we identify the level of group members' role-behaviors; then by using the "whole thread" as the unit of analysis, we recognize the relationship between the current posting, the postings of earlier turns (Chen & Chiu, 2008), and the postings of later turns; and finally, by comparing and contrasting the results of the above two kinds of units of analysis, we interpret our findings in a dialectic way.

Research Method

The Context

487 voluntary participants attended a 6-week inquiry learning in LAIN, in which they were divided into different groups as determined by the topics in which they are interested. Group members in this inquiry activity collectively justify a hypothesis and complete the task.

The members of group C2 were selected to be our research subjects. The C2 group consisted of 6 members whose postings totaled 389 and whose online presence frequency totaled 526 (Table 1). The main topic of the C2 group discussion was "Fog".

Table 1: The frequency of postings and online presence totals of C2 group members.

ID	Posting	Online
010124	150	182
Milkbottle	123	119
Cathyjudy	53	94
Icebox	34	60
Snowlove	28	70
Beer	1	1
	389	526

Since we were interested in how knowledge co-construction developed, we focused on the longer threads, and chose the 15th as the sample discussion thread. The 15th thread consisted of 69 postings which were posted by the 4 members during the second week. Icebox and Beer had not participated in this 15th thread. The long thread provides us a better opportunity to discover the progress and the dynamic interaction in the C2 group discussion.

Data Collection and Analysis

The data log of the discussion forums served as the main resource of data analyses. Data were analyzed to determine each member's participation level and learner-role behavior. In addition to the frequency of total postings and weekly postings, and the timing of the postings; the content of the postings, the interrelationships between them, and the weekly diary of inquiry process were also collected. Two kinds of unit of analysis were used: "single posting" and "the whole thread".

First, the "single posting" was used as the unit of analysis to recognize and identify the level of group members' role-behaviors. We adopted Waters and Gasson's (2006) approach — "the classification of the primary learner-role behaviors" — to examine the contribution of the participant's behavior in the online forum. The eight learner-role behaviors are: Passive-learner, Knowledge-elicitor, Contributor, Vicarious-acknowledger, Closer, Facilitator, Initiator, and Complicator (Table 2). The classification further provided a framework for three levels of involvement: participation, involvement, and social engagement. This classification was accomplished by 1) identifying the contribution of each posting through comparing the relationship of prior postings to the current one in order to recognize the level and the contribution of the posting, and 2) counting the classification of each posting of each group member, and 3) selecting the highest frequency as the most fitting classification to define the participation level of each group member.

Table 2: The classification of primary learner-role behaviors.

level	Form of Behavior	Milkbottle	Cathyjudy	010124	Snowlove
I. Participation	Passive-learner				
	Knowledge-elicitor	1	3		
	Contributor			13 *	
II. Involvement	Vicarious-acknowledger	1	1	1	
	Closer	7		1	
III. Social engagement	Facilitator	19 *	8 *	1	1
	Initiator	2			
	Complicator				3 *
	Total	30	12	16	4

PS. The result of the classification of each member is indicated with *.

From the results showed in Table 2, most of the members fell into the *III. social engagement* level. Milkbottle and Cathyjudy were classified as the *Facilitators* and Snowlove as the *Complicator.* 010124 was the only person to be classified as the *Contributor* and fell into the *I participation* level. It seemed that Snowlove, Milkbottle and Cathyjudy made a higher level of contribution to the group discussion, while 010124 was deemed to have contributed to this group at a relatively superficial level. Interestingly, although 010124 was the most active person in this group in terms of the number of postings, the nature of her postings in thread #15 was more that of a contributor. Table 3 summarized a profile of each member from our online observation notes.

Table 3: Observation of the characteristics of the learners in the C2 group.

ID	Form of behavior	The characteristics
Milkbottle	III *Facilitator*	Actively participates in group discussion, and often integrates other members' opinions into one posting.
Cathyjudy	III *Facilitator*	Actively asks questions, shares her thoughts, and triggers other's participation in the discussion.
010124	I *Contributor*	Actively responds to others with authoritative content via "copy and paste".
Snowlove	III *Complicator*	Contributes fewer postings to the group but is quite aware of the context of the discussion and points out the problems.

In addition to using Waters and Gasson's approach, we purposely reexamined the contribution of each participant by using "the whole thread" as the unit of analysis (Figure 1). In this way, we were able to trace the evolution of the whole 15th thread by referring to both the backward and forward postings and by selecting isolated postings with similar focus within the thread. The time dimension emerged as a pivotal issue to deal with in knowledge co-construction. In order to represent the meaning of every "single posting" in the accomplishment of knowledge co-construction in online discussion, domain experts and learning scientists collaboratively analyzed the 69 postings with these characteristics in mind: 1) the depth of domain knowledge of the posting, 2) the evolution of the concept's physical property, 3) the flow of the group discussion, 4) the interrelationship between the postings, 5) the participation features of the group member, 6) the timing of the posting and the atmosphere of the group discussion, and 7) the nature of the response of group members. The results of the second approach provide a distinctively different picture of group knowledge co-construction from those of the first approach— but these two viewpoints are dialectically interrelated.

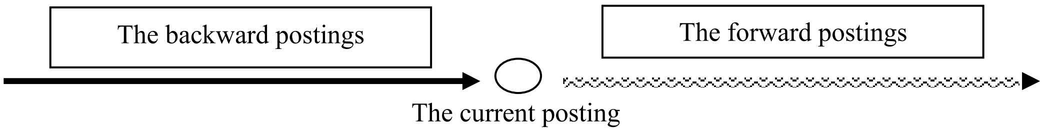


Figure 1. The Analytic Structure for "The Whole Thread" as the Unit of Analysis.

Using the emergence of these particular convergent postings as an index, #21 and #49 were determined to be the watersheds at which to divide the 69 postings into 3 separate pieces for detailed analysis of knowledge co-construction. The 3 separate pieces represented the different meaningful-activities and showed the different phrase existed in the 15^{th} thread. And we found that each posting came from the prior postings and had the effect on the forwards postings. For example, the posting #48 is not just the result of the prior postings #1~#47, but also has the effect on the forward postings #49~#69.

Research Findings

These three episodes contain different stories about how the "whole thread" as the unit of analysis revealed deeper understanding of knowledge co-construction than the "single posting" could. To be specific, the present study demonstrated that the second approach not only interpreted the learning situation of the group discussion differently, but also contrasted sharply with the first approach on the contribution of group members.

Confusion and Hesitation

The first episode shows that the progress of the group discussion had gotten into a confusing condition by using the "single posting" as the unit of analysis. However, using "the whole thread" as the unit of analysis, we discovered that the group members were simply waiting for next opportunity to clarify the problem.

The phenomenon of knowledge co-construction could be a slow process and could have a pre-condition. An authentic situation for knowledge co-construction to develop is in response to encountering a confusing situation. The following excerpt is an example. Milkbottle came up with a hypothesis(#11), Cathyjudy tried to clarify the question(#12), Milkbottle and 010124 not only responded to Cathyjudy but also became mired in confusion thinking about an uncertainty (#13~15).

No	ID	Content
11	Milkbottle	(……)I have an idea about the hypothesis: Fog occurs in the interface of warm and cold air masses. When atmospheric convection is not efficient, the cold and warm air masses come in contact with each other.
12	Cathyjudy	What does it mean that atmospheric convection is not efficient? Does it mean there is no wind??
13	Milkbottle	It might not be possible that there is no wind when cold and warm air contact each other in an air mass within the specific area. I don't know. Can anyone respond to this?
14	010124	No matter whether it is breezy or windy, lower atmosphere is always flowing. The solar energy is the source of all activities in the atmosphere. The air is heated, expanding and drifting upward and the low air occurs at the surface. It is impossible to have no wind because warm air goes up; cool and height-density air replaces the warm air and thereby generates the wind. Therefore it is not possible to happen in a situation with no wind.
16	010124	I don't know if my response was correct. Keep thinking about it.
31	Cathyjudy	The example that the car and the glasses should not be related to the concept of convection.

The appearance of confusion might be invalid in knowledge co-construction. Group members mixed up the concepts of "convection" and "wind". They all felt frustrated and shifted discussion to another topic after #16. Using the "single posting" or "adjacent pair" as unit of analysis, one might consider this episode as ineffective and exclude it from knowledge co-construction. However, using "the whole thread" as the unit of analysis, we discovered that the messy conception had become less unclear. Cathyjudy started to clarify the idea of "convection" in #31, which is far from the earlier turns. That is, linking backwards to earlier turns is as important as forward to later turns in investigating the process of knowledge co-construction.

Furthermore, the discontinuity between smoothness and stagnation would be regarded as an opportunity for evoking growth or new turns. In the first episode, group members appeared to change to another topic because they could not discuss the present topic any further. Later on, we noticed that the group members still cared about the question and kept it in mind. They engaged more and associated relevant ideas to the previous answer. Were we to assume that there was no progress in this episode of group discussion, we might misinterpret their pondering time as valueless in knowledge co-construction. But by putting together pieces of puzzle, that is, related postings of earlier turns as well as of later turns, we discovered that the group members cross from discontinuity and stagnation to renewal of progress as a result of an apparently imperative time period of confusion and hesitation.

From Stick Around to Move Forward

The second episode would appear that the group was engaged in a warm and well-focused atmosphere when using the "single posting" as the unit of analysis. However, using "the whole thread" as the unit of analysis, we found that although the two *Facilitators* actively developed their ideas and presented a sustained discussion, the discussion was not really making further progress and had lapsed into stagnation. Moreover, it was the *Contributor* who eventually advanced the group discussion "from stick around to move forward".

Knowledge co-construction deals with the development of collective and public knowledge. It therefore involves negotiation and a bid for currency of each other's ideas. The ones who contribute more in the discussion forum seem to have a higher level of social engagement, and this results in greater contributions to knowledge co-construction. The interaction between the two facilitators, Cathyjudy and Milkbottle, in the second episode was identified as consisting of effective postings using the "single posting" as the unit of analysis. Interestingly, in some cases "effective discussion" might actually cease growing. In contrast it was the contributor, 010124, who made group progress in knowledge co-construction. Earlier by using the "single posting" as the unit of the analysis in identifying the roles of group members, Cathyjudy and Milkbottle were identified as *Facilitators* who engaged in and contributed more to the discussion, while 010124 was regarded as a *Contributor* who engaged in and contributed less to the discussion because her postings (#37, #38, and #40) mainly consisted of copying and pasting references and never conveyed her own independent opinions.

However, the above interpretation of the roles of group members was overthrown by using "the whole thread" as the unit of analysis in recognizing the roles the members played. Since 010124 came close to Cathyjudy and Milkbottle in the creation and assimilation of knowledge over a long period of time, she paid a great deal of attention to Cathyjudy and Milkbottle's postings, and at the same time, she proceeded to diagnose and justify Cathyjudy and Milkbottle's postings. In the process of following Cathyjudy and Milkbottle's discussion, 010124 seemed to play the role of a semi-outsider, but 010124 participated in and observed the group discussion in a salient fashion. When Cathyjudy and Milkbottle developed a series of 16 postings about the topic of "the relationship between Fog and Wind", 010124 found them to be unfocused. Apparently the 16 postings seemed to go deeper than earlier; 010124 noticed a knowledge gap which stopped the discussion from growing. She therefore tried to speak up and present her opinion. As Lave (1991) claimed from a situated learning perspective: "This is the time that they (members) feel that they belong enough to carry the message"

No.	ID	Content
37 38	010124	Fog occurs more easily in a low-lying area!!! Fog occurs over the land more easily than over the sea.
39	Milkbottle	Low-level cold air should be left behind by the cold front, so it has not enough energy to raise warm air up.
40	010124	Fog takes place at night easier than during the day
41	Milkbottle	010124! Give the reasons of your hypothesis or you will not be able to convince anyone.
42	Milkbottle	Rain in Spring is too light to be felt. Is it a kind of drizzle?
43	010124	Wind at high altitude is stronger than at lower altitude, so I proposed the first hypothesis. Wind over the sea is stronger than over the land, so I proposed the second one. The reason of the third one is that wind is always stronger during the day than at night.
48	Cathyjudy	*According to #15-42* According to 010124, can we summarize them all into a hypothesis? The occurrence of fog is related to wind, because only one key point that 010124 mentioned is wind……

Engagement including inviting and accepting the ideas of others is important in the process of knowledge co-construction. 010124 avoided joining in the discussion about "the relationship between Fog and Wind" in which Cathyjudy and Milkbottle were actively involved. Instead, 010124 subtlely joined them by initiating a title about "hypotheses involving the factor of fog" (#37, #38 and #40). Although 010124's 3 postings were mainly authoritative information forwarded from websites, these 3 postings did evoke negotiation within the group.

The subtlety of the value of a posting is not dependent upon its having the appearance of being a "copy and paste" contribution, but on its timing and the need it filled during knowledge co-construction. Using the "single posting" approach, postings such as "copy and paste" are defined as a superficial contributions. However, using "the whole thread" approach, the contribution of such postings changed. In our case, Milkbottle noticed these 3 information-based postings, and negotiated a bid for them (Wenger, 1998) (#41). Then, Milkbottle invited 010124 to elaborate more upon her opinion. As a result of Milkbottle's invitation, 010124's posting was able to increase in value and become used as a bargaining chip, adopted by other group members (Zhou & Stahl, 2008). At this moment, 010124's long engagement did clearly produce the positive effect of acquiring more knowledge. 010124 used the knowledge that was learned from tracking the history of the evolution of the group discussion to validate and support her opinion (#43). Then, Cathyjudy and Milkbottle adopted 010124's opinion (#48) and it became legitimate knowledge in the group discussion. In the above mentioned case, we can see that Cathyjudy and Milkbottle's willing adoption created an opportunity for 010124, who played a crucial role in bringing about opportunities for group success. 010124 took advantage of an opportunity to justify and produce a dramatic shift which improved group discussion, as a result of which, Cathyjudy and Milkbottle made an adjustment which led the group discussion toward greater progress and away from its stymied state – thus expanding the lifespan and longevity of threads. In this second episode, we can see and understand how groups maintain their continuity of interaction across discontinuities.

Doing Justification and Breaking Through

The third episode showed that the *Complicator*'s opinion was more powerful and the *Contributor*'s posting weaker in the group discussion when using the "single posting" as the unit of analysis. However, when using "the whole thread" as the unit of analysis, we found that nobody but the *Contributor* cared about the *Complicator*'s opinion. Again, the *Contributor* justified and supported the correctness of the *Complicator*'s opinion by copying and pasting. In this process of justification, the *Contributor* clarified misconceptions in the group discussion, and affected the direction of the group discussion.

The *Contributor* who was regarded as demonstrating a lower level of contribution, may play a critical role with the *Complicator* in knowledge co-construction. In the third episode, group members seemed to exhibit a great deal of commonality and consensus; but the condition was actually somewhat ambiguous at this stage.

The concept of "wind" as the main hypothesis(#43) was started back in #18, and went through 25 postings (#18~43), eventually to be accepted and integrated in the group hypothesis for the group member who had engaged more over a long period of time. But, the *Complicator*, Snowlove, who always remained silent in the group, at this stage broke her silence and submitted 3 postings (#45~47) to address the question of the concept of "wind" as the main hypothesis and one that was in her opinion justified, specifically "advection fog is not restricted by wind speed, but the advection of warm and wet air.……". When we identified Snowlove's 3 postings using "single posting" as the unit of analysis, we had considered them as deeper and more contributive to the group discussion. However, we found that the 3 postings had different meaning-making for other members in the process of discussion when using "the whole thread" as unit of our analysis. The 3 postings' meaning for other members is as follows.

No.	ID	Content
45 46 47	Snowlove	This is used data...... I thought radiation fog took place on land mostly and the weather must be less wind. …… advection fog is not restricted by wind speed, but the advection of warm and wet air.…… Maybe other teams study radiation fog and advection fog, so can we start from another point of view, like looking for some materials of other kinds of fog, then discussing?
48	Cathyjudy	*According to #15-42.* According to 010124, can we summarize all into a hypothesis? The occurrence of fog is related to wind, because of only one key point that 010124 mentioned is wind……
49~50	Milkbottle	The conditions of formation of fog: 1. low temperature, 2 high humidity, 3. less wind, and condensation nuclei.
51, 52, 55, 57~60	010124	*frontal fog, fog, fog, Upslope fog, Spring fog in early morning is a sign of clear sky. Summer fog in early morning, Heavy rain in the afternoon* (All the contents forwarded from the website were omitted due to the limited space)
53	Milkbottle	*Frontal fog, fog.*
62	Cathyjudy	*The main hypothesis.* When the cold and warm air contact each other, fogs take place at where the large temperature difference, more humid warm air, and more condensation nuclei. I specially point out the humidity of warm air…

65 67	Milkbottle	*The main hypothesis.* In nature, cold air is below warm air when cold and warm air contact each other. A big temperature difference means the difference between the temperature of cold air and warm air. More temperature difference, suitable wind speed, higher humidity and more nuclei in the air, then fog forms easily. …… *The main hypothesis.* Humidity of warm air
68	Cathyjudy	*The explanation for the main variable.* The main variable is not just humidity, but the humidity of warm air. If the humidity of warm air is not high enough and the saturated water vapor pressure becomes lower, it does not reach saturation.

Although the *Complicator,* Snowlove's opinion was strong (#45~47), the group members had given it an unfriendly reception, in contrast to that which they gave to the *Contributor,* 010124. We could see a decrease in postings and a silent and conflicted atmosphere when Snowlove attempted to raise a point counter to the direction of the current group discussion. The *Faciliator*, Cathyjudy, tried to counter Snowlove's opinion by using 010124's previous opinion (#48). The *Faciliator* Milkbottle kept summarizing group members' opinions but excluded Snowlove's proposal from the "the summarization of hypotheses postings" and did not respond to Snowlove. At this point, Snowlove's posting was ignored by the group members.

We recognize a bid to negotiate interactively between the *facilitator* and the *Contributor*: the *Contributor* 010124 justified the *Complicator,* Snowlove's, posting to help build a bridge between their conflicting opinions. It appeared that nobody cared for Snowlove's argument (#45~47*)*, but 010124 did pay sufficient attention to it. Instead of using personal internal resources to judge the opinions of other group members, 010124 sought an external resource to evaluate Snowlove's claim. 010124 took Snowlove's opinion and looked for authoritative references to check it, then copied and pasted seven authoritative references (#51, 52, 55, 57~60) to justify Snowlove's opinion as right. In this way, 01024's posting eventually evoked negotiation within group. The group members renewed negotiation of the value of Snowlove's claim. Milkbottle also sought out the external resource to justify Snowlove's posting (#53) and then, convinced of its validity, supported Snowlove's argument. Snowlove's argument became legitimate knowledge for the group members. The *Facilitators*, Cathyjudy and Milkbottle, adopted the *Complicator,* Snowlove's, opinion to make an adjustment to the direction of group discussion (#62~68). We could see that the vindication of this view revived the progress of group discussion which then went forward again toward knowledge co-construction.

Discussion and Conclusion

Our study introduces new ideas for knowledge co-construction. First, previous research explored the frequent "give and take" of the success dialogue of episodes. The present study shows that the occurrence of knowledge co-construction may start right in the process of confusion when "give and take" goes less smoothly among the group members. If we don't recognize the "confusion and hesitation" stage of group discussion, we may not capture and recognize that starting moment and arrive at a complete picture of knowledge co-construction.

Second, most related research has focused on the predominant group members, such as the leaders or the active ones, and has tried to capture the learning process from their dialogue interaction. In the present study, we uncover the complex dynamics of group practice in ways that might further illuminate the phenomenon that weaker and less active members do bring about a turning point in knowledge co-construction. If we had not focused equally on the role of ordinary people in the group discussion, we would have achieved merely a partial understanding of the knowledge co-construction.

Insights into knowledge co-construction are developed by re-considering the analytic method for group discussion. We compare and contrast the results of analysis by using the "single posting" and "the whole thread" as the unit of analysis. We first adopted the "single posting" as the unit of analysis for recognizing the content of each posting, classified each posting, and identified the classification of the primary learner-role behaviors. Then, we adopted "the whole thread" as the unit of analysis for interpreting the meaning of the current posting by carefully investigating both the earlier and the later ones. Using this compare-and-contrast method, we demonstrated that most of the meaning differed from that determined before, and that the classification of primary learner-role behaviors was modified.

As with any other research, certain limitations of our findings must be noted. First, the sampling of the group. Lain was a naturalistic virtual setting in which participants in this study were able to withdraw, intermit, wander, or engage at will and at their own pace. C2 was selected from the 64 groups because the members had more focused and on-topic discussion, and members participated more equally in the discussion. We considered knowledge co-construction to be more recognizable in this group. The findings therefore need to be viewed with caution as they would not be representative of other types of online group collaboration. Second, the sampling of the thread. In analyzing co-construction among group members, we purposely selected the thread with the longest lifespan, the 15th thread. Future studies with short threads would be useful to verify our findings.

The results of this study have implications for analytic methods to examine knowledge co-construction. In conventional analytic methods, previous research ignored the significance of the time dimension in group discussion when quantifying data in classifying the role of each group member and in recognizing the relationships between group members. In contrast to the conventional view, we argue that we must necessarily focus on contextually-situated processes and the time dimension of the process of knowledge co-construction, and must avoid making unwarranted assumptions about the role of group members. We suggest using "the whole thread" as the unit of analysis in tracking critical moments and identifying the contribution of members in knowledge co-construction. From "the whole thread as the unit of analysis" perspective, we further suggest that both linking backwards to earlier turns and linking forward to later turns were equally important in determining the meaning of a posting, and result in a more complete picture of knowledge co-construction.

References

Aalst, J.V. (2009). Distinguishing knowledge-sharing, knowledge-construction, and knowledge-creation discourses. *The International Journal of Computer-Supported Collaborative Learning (ijCSCL), 4* (3), 259-287.

Chen, G., & Chiu, M.M. (2008). Online discussion processes: Effects of earlier messages' evaluations, knowledge content, social cues and personal information on later messages. *Computers & Education*, *50*(3), 678-692.

Chiu, M.M. (2008). Flowing toward correct contributions during group problem solving: A statistical discourse analysis. *The Journal of the Learning Sciences, 17*(3), 415–463.

Dennen, P.V., & Wieland, K. (2007). From interaction to intersubjectivity: Facilitating online group discourse processes. *Distance Education*, *28*(3), 281-297.

Diggelen, W.V., Janssen, J., & Overdijk, M. (2008). *Analyzing and Presenting Interaction Data: A Teacher, Student and Researcher Perspective.* Paper presented at the International Conference of the Learning Sciences (ICLS), Utrecht, Netherlands.

Hewitt, J. (2005). Toward an understanding of how threads die in asynchronous computer conferences. *Journal of the Learning Sciences, 14*(4), 567-589.

Hmelo-Silver, C.E., Chernobilsky, E., & Mastov, O. (2006). *Representations for analyzing tool-mediated collaborative learning.* Paper presented at the 7th International Conference of the Learning Sciences, Mahwah, NJ.

Jeong, H., & Chi, M.T.H. (2007). Knowledge convergence during collaborative learning. *Instructional Science*, 35, 287-315.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.

Reimann, P. (2009). Time is precious: Variable- and event-centred approaches to process analysis in CSCL research. *International Journal of Computer-supported Collaborative Learning*, *4*(3), 239-257.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal education in a knowledge society* (pp. 67-98). Chicago: Open Court.

Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In K. Sawyer (Eds.), *Cambridge handbook of the learning sciences* (pp. 97-118). New York: Cambridge University Press.

Schrire, S. (2006). Knowledge building in asynchronous discussion groups: Going beyond quantitative analysis. *Computers & Education, 46*, 49–70.

Stahl, G. (2008). Chat on collaborative knowledge building. *QWERTY, 3*(1), 67-78.

Stahl, G. (2009). *Studying virtual math teams.* New York, NY: Springer.

Stahl, G. (2010). Social practices of group cognition in virtual math teams. In S. Ludvigsen, A. Lund, & R. Saljo (Eds.), *Learning in social practices. ICT and new artifacts: Transformation of social and cultural practices.* London: Pergamon.

Stahl, G. (2010). How I view learning and thinking in CSCL groups. *Research and Practice in Technology Enhanced Learning (RPTEL),* 1-23.

Suthers, D.D. (2006). Technology affordances for intersubjective meaning-making: A research agenda for CSCL. *International Journal of Computers Supported Collaborative Learning, 1*(3), 315-337.

Suthers, D.D., Dwyer, N., Vatrapu, R., & Medina, R. (2007). *An abstract transcript notation for analyzing interactional construction of meaning in online learning.* Paper presented at the 40th Hawai`i International Conference on the System Sciences (HICSS-34),Waikoloa, Hawai`i.

Waters, J., & Gasson, S. (2006). *Social Engagement in an Online Community of Inquiry.* Paper presented at the 27th International Conference on Information Systems (ICIS 2006), Milwaukee, WI.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity.* Cambridge: University Press.

Zhou, N. (2009). Question co-construction in VMT chats. In G., Stahl (Ed.). *Studying Virtual Math Teams.* New York, NY: Springer Press.

Acknowledgements

This work was supported, in part, by National Science Counsel Grants NSC 97-2511-S-008 -002 -MY3 and NSC 99-2631-S-008-001-.

Towards an Understanding of "Listening" in Online Discussions: A Cluster Analysis of Learners' Interaction Patterns

Alyssa Friend Wise, Farshid Marbouti, Jennifer Speer, Ying-Ting Hsiao, Simon Fraser University, 250-13450 102 Avenue, Surrey BC, Canada, V3T 0A3
Email: afw3@sfu.ca, fmarbout@sfu.ca, jspeer@sfu.ca, yha73@sfu.ca

Abstract: Conducting learning conversations through online discussion forums differs from face-to-face conversations as learners can be selective in what comments they choose to "listen" to, when they chose to do so, and how long they spend attending to them. Using cluster analysis of learners' click-stream data from an online discussion forum, this study identified three patterns of behaviors that differentiate between learners: Superficial Listeners, Intermittent Talkers; Concentrated Listeners, Integrated Talkers; and Broad Listeners, Reflective Talkers. Clusters differed in the amount of time spent listening, number of posts attended to, number and length of their sessions, and number of own posts contributed and reviewed. Clusters did not differ in percentage of posts scanned (vs. read), time to compose posts, length of posts made, or final course grades. Results are compared to interaction styles found for other online tools and implications for practice and future research are discussed.

Introduction

Conversation conducted through online discussion forums continues to gain interest as a vehicle for learning (e.g. Barab, Kling & Gray, 2004; Luppicini, 2007). Theoretically, discussion forums can support students in actively engaging with others to build understanding as they externalize their thoughts, hear alternative points of view and work collectively to negotiate meaning (Boulos & Wheeler, 2007). Such conversations can thus support knowledge construction both from the perspective of a group collectively improving ideas and from the perspective of the individuals in the group deepening their personal understanding (Stahl, 2003). However, online discussions don't always live up to this promise; in practice it is common to find fractured and incoherent conversations (Herring, 1999) with low levels of interactivity between students (Thomas, 2002).

Theoretical Framework - Speaking and Listening as Core Elements of Conversation

In the model above, knowledge construction occurs as learners share the ideas and work with the ideas of others. Similar to face-to-face conversation this can be thought of as "speaking" (externalizing one's ideas) and "listening" (taking in the externalizations of others). But compared to face-to-face conversations, in online discussions learners have greater control of what comments they choose to attend to, when they chose to do so, and how long they spend "listening" to them. The majority of research on collaboration in online discussions has focused on the "speaking" aspects of conversation – studying the messages learners contribute and how they interact to produce group meaning-making (e.g. Arvaja, 2007; De Wever et al., 2006). Much less attention has been paid to the interactions learners have with existing messages and the often extended sequences of listening behaviors they engage in during the processes leading up to making a contribution (e.g. how learners navigate the existing discussion; how many and which messages they choose to open; how they interact with this content once opened; how they compose their own contributions). These "online listening behaviors" are core to interactivity in online discussions and are an important part of the knowledge construction process that can influence both the contributions made and the uptake of ideas between learners (Suthers, 2006).

Initial work suggests that for many learners, their interactions with previous messages are brief and superficial. Thomas (2002) found that on average students read a low number of messages compared to the number they posted, and a substantial portion of messages did not meaningfully refer to previous ones. Hewitt (2003; 2005) found that while most students did read at least one message before composing their own, they took a single-pass strategy that focused almost exclusively on the most recently posted messages. Learner listening behaviors appear to be sub-optimal for learning (Thomas, 2002) and heavily influenced by the interface in which they take place (Kear, 2001). While past studies provide a broad outline of the situation, they do not give detailed information about how particular students interact with the discussions. Nor do they help us identify productive interaction strategies which might be encouraged in other students. Further research on how learners listen in online spaces and their motivations for doing so is needed to build a more complete picture of the process of learning through online conversation and create a theoretical foundation for designing interventions to support more effective listening behaviors.

The Current Study

This work builds on previous efforts by examining in fine-grained detail what online listening behaviors particular learners engage in. Specifically, the goal of this study was to identify and group particular sets of behaviors that characterize different interaction patterns learners use to interact with the existing posts in discussion forums. In future work, these styles can be classified as more or less productive and used to help design interventions to support more effective listening behaviors in all students.

Cluster Analysis

Cluster analysis is a technique used to identify naturally occurring groups that has productively been used to better understand learner behaviors in a variety of digital spaces. For example, in Barab, Bowdish, and Lawless' (1997) work looking at the use of a hypermedia computer kiosk, four types of user activity patterns were identified: model users (motivated to solve problems, least number of navigation choices, few deviations from task path), disenchanted volunteers (explored very little, low number of choice, time spent on screens), feature explorers (feature oriented, many navigation choices, used help screens), and cyber cartographers (goal-directed, fewer navigation choices but longest time spent on each screen). In their work looking at activity in a learner-centered online distance education environment, del Valle & Duffy (2007) found that learners could be characterized as mastery-oriented (high number of sessions, activities and resource use), task-focused (shorter more concentrated work time but similar amount of activities and resource use), or minimalists (few sessions, activities and resources spread over a long time). In the current study, cluster analysis is used to look for patterns in online listening behaviors that constitute different styles of interaction.

Research Questions

1. What are the distinctive approaches learners take in interacting with existing comments in asynchronous online discussion forums?
2. What are the characteristics of these approaches with respect to "listening" and "speaking" in online discussions?

Methods

Participants

Ninety-six of 113 students enrolled in a blended (face-to-face and online) undergraduate business course on organizational behavior at a mid-sized Canadian university agreed to participate in this study. Participants were evenly divided on gender (51% female) with the majority in the first two years of their study (85%) and 22 years old or younger (94%). Less than a third of participants were native English speakers (28%), which may not be representative of a typical western university classroom. Final grade distribution for the class was typical for the university with a mean in the B range.

Learning Environment and Discussion Task

The face-to-face component of the course consisted of a two-hour weekly whole class lecture and one of five one-hour weekly tutorial sessions conducted in groups of 20-23 students. For the online component, each tutorial was split into two sub-groups to participate in three asynchronous discussions worth 9% of their grade. Discussions were a week long and ran Saturday to Friday. In the first week, all sub-groups took part in an ungraded "Introductions" discussion. For the next six weeks, the two sub-groups alternated discussion weeks (one sub-group participated in discussions one, three, and five, the other in discussions two, four, and six) though all discussions were available to both sub-groups throughout the six week period. Discussions took place in Phorum, a basic asynchronous threaded discussion forum. In each discussion, students were asked to collectively solve an authentic organizational behaviour challenge. Students were required to be actively involved in the discussions: contribute more than once and make comments to progress the group's discussion. One student was chosen in tutorial at the end of each assignment to summarize the challenge and solution orally.

Data Extraction and Processing

Participants' click stream data was collected based on their activity in the discussion forum. Every time a student clicked to read, create, or edit a post in the discussion forum, a data entry was created logging the action taken, the identity and length of the post, and a time stamp. Data was collected from the all the organizational behavior challenge discussions; data was collected for the Introductions forum only if it was generated after the challenges were in progress. Extracted data was filtered by participants' user ID. Actions were coded as either "views" (opening others' posts), "posts" (creating posts), "reviews" (revisiting one's own posts later), or "edits" (making changes to one's previously submitted posts). Time between subsequent actions was subtracted to calculate the action duration. Views were then further categorized as either scans or reads based on the speed with which the post was viewed. A threshold of 6.5 words per second (wps, calculated as the ratio of the time spent viewing a post to the word length of the post viewed), was used based on a maximum reading speed of 6.39 wps for online messages found by Hewitt, Brett, and Peters (2007). Only views that fell below this speed were categorized as reads, other views were categorized as scans.

Because users did not have to formally log-out of the system, action length needed to be manually divided into sessions of use (e.g. if a learner reads a post today and then another one tomorrow, the length of the first read action would be calculated as over 12 hours). Following the precedent set by del Valle and Duffy (2007), we determined a maximum threshold of allowed action length. Frequency tables showed that 88.8% of all actions and 96.9% of post actions (typically longer and less likely to be abandoned in progress) took 60 minutes or less, with a sharp drop in frequency after this point. Thus 60 minutes was taken to be the maximum allowed action length. Actions exceeding this were determined to be the end of a session and their calculated duration was replaced with an average for the action, taking into account the length of the post read or created and the mean reading or posting speed for that learner.

Variables

Seven variables were judged as best representing the different facets of students' interactions in the discussions. These variables reflect how students chose to visit the discussions, the breadth and depth of listening they engaged in while there, their level of speaking in the discussion and the degree of integration with their listening behaviors, and finally the degree to which they attended to their own voice in the forum. In addition to the variables used in the cluster analysis, six other variables were used to further investigate the behavior of cluster members and characterize the differences (and similarities) between clusters.

Cluster Analysis Variables

Average Length of Session served as a measure of the degree to which students spent continuous periods of time in the discussion. It was calculated as the total time (in minutes) the student spent in the system divided by the number of sessions in which the student used the tool.

Percent of Posts Viewed at Least Once served as a measure of the breadth with which learners listened to others in the discussion. It was calculated as the number of unique posts (made by others) that a student opened divided by the total number of posts made by their classmates to the discussion forum.

Percent of Total Views that were Reads (Not Scans) served as an initial measure of the depth with which learners listened to others in the discussion. It was calculated as the number of views of others' posts that were read at a rate higher than 6.5 words per second divided by the total number of views made.

Average Length of Time Reading a Post served as a further measure of the depth with which learners listened to others in the discussion**.** It was calculated based on posts read (not scanned) as the total time others' posts were open in minutes, divided by the number of reads.

Average Number of Posts Contributed per Assignment served as a measure of the quantity of student's speaking. It was calculated as the total number of posts made divided by the number of assignments.

Percent of Sessions with Posting Actions served as a measure of the degree to which students integrated their listening and speaking behaviors. It was calculated as the number of sessions in which a post was made divided by the total number of sessions a student had.

Average Number of Reviews per Assignment served as a measure of the degree to which learners listened to their own speaking. It was calculated as the total number of times a learner opened their own posts (not immediately following their creation) divided by the number of assignments.

Additional Comparison Variables

Average Number of Sessions per Assignment served as a measure of the degree to which learners concentrated or distributed their visits to the discussion forums**.** It was calculated as the total number of sessions divided by the number of assignments.

Average Number of Views per Assignment served as an additional measure of the breadth with which learners listened to others in the discussion. Calculated as total number of posts (written by others) that were opened divided by the number of assignments.

Average Number of Reads before Contributing a Post served as an additional measure of the integration of listening and speaking behaviors. It was calculated as the total number of others' posts read before making a post in a given session divided by the number of posts made.

Average Number of Words per Post served as a measure of the quantity of speaking a student did in a discussion. It was calculated as the total number of words posted divided by the total number of posts created.

Average Length of Time Creating a Post served as a measure of the care learners put into their speaking within the discussion forum. It was calculated as the total time the Creating a New Post window was open in minutes divided by the total number of posts created.

Final Grade served as a measure of the overall level of performance in the course. It was calculated by the instructor based on the discussion assignment, in-class quizzes, and a midterm and final exam.

Cluster Analysis

There are no rules-of-thumb about the sample size necessary for cluster analysis, since this depends on how participants are distributed across the variable space. In linear regression, a standard guideline is a ratio of 10-20 cases per variable; however, cluster analyses have been successful with ratios as low as 7.4 (del Valle & Duffy, 2007). One of the 96 cases was removed from analysis because it appeared as an outlier on multiple measures. Thus, in this study, cluster analysis was used to group 95 learners based on seven variables, producing a ratio of 13.6 cases per variable, well within acceptable limits.

To examine within-subject activity profiles, Ward's (1963) hierarchical clustering technique and the square Euclidean distance model were used to determine the distance between clusters. All scores in the cluster analysis were standardized to account for the differing scales or measurement. Following the procedures of Barab et al. (1997), a scree plot was created to evaluate the between-cluster differences. Visual examination was used to determine the point at which the plot leveled off, indicating that a clustering solution after this point would not have meaningful differences between the additional groups. A series of ANOVAs examining between-group differences were conducted to confirm that a quality solution was obtained. Tukey post-hoc comparisons with a Bonferroni correction of the alpha level were used to detect significant differences between the clusters. ANOVA and post-hoc analyses were also performed on the additional six variables of interest.

Results

Descriptive Statistics

In total the 95 participants performed 17,695 actions in the system with an average of 186 (SD=127) actions per learner. Each learner engaged with the discussion on average 22 (SD=16) times (sessions) or about eight times for each discussion. For each discussion, learners viewed an average of 53 (SD=37) posts and created 2.2 of their own (SD=.95). Posts averaged 192 words (SD=74) and overall 51% of viewed posts were actually read (not scanned) (SD=13%). Of the almost three hours (171 minutes, SD=136) that learners spent in each discussion, on average, 73% (125 minutes, SD=113) were spent in listening activities. Of the remaining time, 18% (30 minutes; SD=21) was spent creating posts and 6% (10 minutes; SD=20) was spent on reviewing their own posts. The remaining 3% was attributed to automatic system-generated actions. The high standard deviations indicate great variety in the degree to which individuals interacted with the discussion forum further indicating the need for the more fine-grained analysis of individual behavior that follows.

Cluster Analysis

The scree plot for the data showed leveling between clusters 3 and 4 (see Figure 1), suggesting that a three cluster solution is the best fit for this data. Data was resorted and analysis repeated to test the robustness of the solution; cluster membership did not change. ANOVAs examining showed significant differences between clusters on six of the seven grouping variables, confirming a quality solution. There was no discernible pattern of alignment between clusters and discussion groups, all discussion groups had members from at least two different clusters and eight of the ten discussion groups had members from all three clusters. The characteristics of each of the three clusters are described in detail below and summarized in Table 1. Clusters whose levels of a variable could not be distinguished from each other in the post-hoc tests share one or more subsets in common; clusters which had significantly different levels for a variable have no subsets in common.

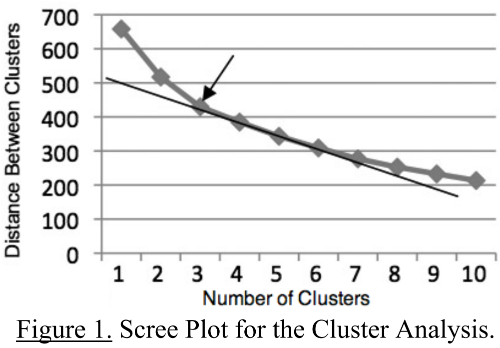


Figure 1. Scree Plot for the Cluster Analysis.

Cluster 1: Superficial Listeners, Intermittent Talkers

The first cluster accounted for 31% of the participants (n=29). In contrast to the other clusters, these learners had a short average session length (14 min), viewed only a moderate amount of their classmates' posts (on average 65%), read (as opposed to scanned) less than 50% of these, and spent an average of just 3 minutes per post on the posts they did read. Looking at the short length of both sessions and post reads, these learners can be characterized as Superficial Listeners. Learners in Cluster 1 created an average of less than two posts per assignment and performed infrequent reviews of these. While both the number of posts made and reviews conducted were significantly less than Cluster 3, the percent of sessions in which they made posts did not differ significantly. Thus a large percentage of Cluster 1 members' sessions (almost 80%) were "listening only" sessions. This repeated listening only behaviour, in combination with the low number of posts made, suggests that the members of this cluster can be characterized as Intermittent Talkers.

Table 1: Results of cluster analysis and between cluster comparisons.

		Cluster 1 *Superficial Listeners, Intermittent Talkers* n = 29	Cluster 2 *Concentrated Listeners, Integrated Talkers* n = 47	Cluster 3 *Broad Listeners, Reflective Talkers* n = 19	F (2,92)
Average Length Of Session	Subset 1		30.02 (9.13)	26.94 (8.48)	37.15*
	Subset 2	13.92 (5.37)			
Percent of Posts Viewed at Least Once	Subset 1			0.96 (0.06)	39.82*
	Subset 2	0.65 (0.20)	0.52 (0.20)		
Percent of Total Views that were Reads (not Scans)		0.45 (0.14)	0.54 (0.13)	0.51 (0.08)	4.67
Average Length of Time Reading a Post	Subset 1		5.18 (2.26)	4.63 (1.77)	9.98*
	Subset 2	3.15 (1.38)		4.63 (1.77)	
Average Number of Posts Contributed per Assignment	Subset 1			3.04 (0.68)	14.88*
	Subset 2	1.69 (0.65)	2.18 (0.98)		
Percent of Sessions with Posting Actions	Subset 1		0.40 (0.16)		25.84*
	Subset 2	0.23 (0.11)		0.18 (0.06)	
Average Number of Reviews per Assignment	Subset 1			10.16 (7.22)	38.29*
	Subset 2	2.09 (1.64)	2.09 (1.95)		

* $p < .0038$ (.05/13)

Cluster 2: Concentrated Listeners, Integrated Talkers

The second cluster consisted of the largest number of learners, 49% of all participants (n=47). Learners in this cluster were Concentrated Listeners; they attended to a smaller percentage of their classmate's posts (52%) than members of Cluster 3 (Broad Listeners), but for the posts they did read they spent significantly longer (5.18 min on average) than members of Cluster 1 (Superficial Listeners). In addition, their average session length was 30 min, equivalent to that of Cluster 3 (Broad Listeners) and significantly greater than the 14 min average for Cluster 1 (Superficial Listeners). Cluster 2 members were also distinct in the significantly higher proportion of sessions in which they contributed to the discussion (40%). This implies that more of their listening activities occurred in sessions in which they also posted and that they had fewer "listening only" sessions. In this sense they were Integrated Talkers, integrating their posting with their listening more than members of the other clusters. While integrated in their talking, Cluster 2 learners were not prolific, contributing an average of just over two posts per assignment, equivalent to the contribution rate for the Cluster 1 (Intermittent Talkers) and significantly less than the rate for Cluster 3 (Reflective Talkers).

Cluster 3: Broad Listeners, Reflective Talkers

The final cluster consisted of a small group of learners, making up 20% of the total study sample (n=19). The most striking characteristic of this cluster is the extremely large proportion of the discussion that they listened to in some way. On average members of this cluster viewed 96% of all the posts contributed by their groupmates at least once. This is substantially more than the 52-65% of unique posts viewed by learners in the other two clusters. Cluster 3 learners enacted their participation in sessions that lasted almost a half hour; this is similar to learners in Cluster 2 (Concentrated Listeners), but almost double that of Cluster 1 (Superficial Listeners). Additionally, a large percentage of their sessions (over 80%) were "listening only" sessions in which they viewed classmates' posts, but did not contribute any of their own. Thus, learners in this cluster can be characterizes as Broad Listeners who seemed to put considerable emphasis on listening and interacted in some way with a high proportion of the discussion comments contributed. Interestingly, these learners did not differ from the other clusters in how they interacted with the posts once they were open; 51% of the posts that they viewed were only scanned, and for the posts they did read, the average length of time spent on a post (4.63 minutes) did not differ significantly from either other cluster. Learners in Cluster 3 were also prolific talkers, contributing a significantly higher number of posts for each assignment than the other clusters. Interestingly, they also focused considerably on listening to their own voice—on average Cluster 3 learners revisited their own contributions ten times per assignment, an average of over three visits for each post made. This is significantly greater than the number of reviews by learners in the other clusters and suggests that these learners may be engaging in some reflective activities. Thus we characterize Cluster 3 members as Reflective Talkers.

Additional Comparisons

Cluster 3 (Broad Listeners) showed significantly higher levels of participation than both other clusters on two of the six additional variables examined: Average Number of Sessions per Assignment and Average Number of Views per Assignment (see Table 2). Broad Listeners had almost twice as many sessions as members of Cluster 1 (Superficial Listeners) and almost three times as many sessions as members of Cluster 2 (Concentrated Listeners). This indicates that they achieved their increased listening not only through extended sessions (see earlier results) but also through engaging in a larger number of sessions. Broad Listeners also had two and a half times as many total views as members of the other two clusters; this is an even higher ratio than the difference in the percentage of posts they viewed at least once, indicating that another difference between the clusters is the frequency with which they re-viewed posts.

The three clusters did not differ significantly on the other four additional variables examined: Average Number of Reads Before Creating a Post, Average Number of Words per Post, Average Length of Time Creating a Post, and Final Grade. The finding that learners in all three clusters read around the same number of posts (three to seven) before making their own suggests that member of Cluster 3 (Broad Listeners) did not achieve their additional listening through more extended listen-then-speak chains, but rather through additional cycles of listen-then-speak and listening-only sessions. As well, the similarity in the average length of posts (175 – 210 words) and time spent creating them (12-16 min) indicates that members of Cluster 3 achieved their additional speaking through supplemental posts, not through longer or more thought-out ones. Finally, the lack of significant difference between the clusters for the final course grade indicates that learners with all three styles performed equally well in the course overall.

Table 2: Results of additional between cluster comparisons.

		Cluster 1 *Superficial Listeners, Intermittent Talkers* n = 29	Cluster 2 *Concentrated Listeners, Integrated Talkers* n = 47	Cluster 3 *Broad Listeners, Reflective Talkers* n = 19	F (2,92)
Average Number of Sessions per Assignment	Subset 1			13.56 (5.48)	27.55*
	Subset 2	7.49 (3.42)	4.77 (4.37)		
Average Number of Views per Assignment	Subset 1			109.19 (29.44)	70.15*
	Subset 2	44.10 (20.53)	35.44 (22.34)		
Average Number of Reads before Contributing a Post		3.60 (2.38)	5.16 (4.48)	7.03 (2.61)	5.17
Average Number of Words per Post		207.91 (72.00)	186.99 (83.61)	178.79 (47.66)	1.07
Average Length of Time Creating a Post		12.61 (8.23)	13.16 (7.33)	15.50 (5.60)	0.97
Final Grade		75.77 (6.95)	72.81 (10.97)	78.96 (5.12)	3.38

* $p < .0038$ (.05/13)

Discussion

The goal of this study was to understand how learners "listen" in online discussions. Three distinct patterns were identified: Superficial Listeners, Intermittent Talkers; Concentrated Listeners, Focused Talkers; and Broad Listeners, Reflective Talkers. Below we contextualize these findings in past research on online interaction styles and discuss their implications for research and practice.

Comparisons with Past Research, Implications for Future Research

At one extreme, Cluster 1 "Superficial Listeners, Intermittent Talkers" had perfunctory engagement in the online discussions, spending little time listening to others, and talking sporadically. This behavior resembles that of Barab et al.'s (1997) "Disenchanted Volunteers" who showed little interest in exploring more information than necessary. The Superficial Listeners also show a similar pattern to de Valle and Duffy's (2007) "Minimalist" cluster who tried to fulfill their course requirements with a minimum of effort. However, contrary to their context where minimalist learners could protract the duration taken to complete their work, our

Superficial Listeners were forced to participate within the timeline of a week-long discussion. Thus even though they divided their participation into an average of seven short sessions per discussion, these were concentrated within a single week, potentially leading to a more coherent participation experience. Similar to del Valle and Duffy's finding that minimalists managed to complete their course successfully with no difference in the quality of their final product, our Superficial Listeners showed an equal level of performance in their course grade with the other clusters. This may present evidence for a blended context of what del Valle (2006, p116) notes is well known to practitioners but not often discussed in the literature: "the reality of a number of [distance education] learners who focus on course completion with a low commitment that results not in attrition, but in course completions with the minimum work possible." It may be that in this case, Superficial Listeners' low level of interaction with the discussion was enough to develop a sufficient understanding of the ideas being discussed. Another alternative is that the instructor of the course did not make the content of the discussions forums central to the course and its evaluation. In future research, we will work in a more controlled context to examine whether listening in this superficial way impacts what individuals learn from and contribute to online discussions. We will also test whether this stems from a work-avoidance approach to their online discussions and if so, examine their motivations for this.

Cluster 2, "Concentrated Listeners, Integrated Talkers" interacted with the discussion forum for a similar amount of time overall as the Superficial Listeners, but with longer sessions than this group and fewer session than the Broad Listeners. Thus, their participation was condensed, focusing on listening to particular parts of the conversation in a limited number of extended sessions. del Valle & Duffy (2007) characterize similar behaviors seen in their learners as "task-oriented" – focused on fulfilling requirements expeditiously. However while their task-focused learners used a similar number of learning resources to their mastery learners, our Concentrated Listeners listened to fewer comments from their peers than the Broad Listeners. Interestingly, similar to our results, del Valle & Duffy found that their task-focused learners performed as well on the final performance measure for the course as learners in the mastery cluster and Barab et al. (1997) refer to learners with performance oriented behaviors as "model users." This suggests that a task-focused orientation is not necessarily a bad thing; future research is needed to determine whether the concentration of participation and number of messages listened to impacts individual learning in a discussion or the quality of comments made.

Finally, Cluster 3, "Broad Listeners, Reflective Talkers" were high volume participants who spent the most time in the system, with a large numbers of sessions and listening events per assignment, as well as posts and reviews. Compared with the clusters found by del Valle & Duffy (2007), these learners seem quite similar to their "mastery-oriented" group (high amount of sessions, time, and learning resources used). However, they differed from Barab et al.'s (1997) mastery group (called "cyber cartographers") who spent more time on fewer navigation choices. Our learners showed increased breath of participation (rather than depth); they viewed many more posts, but spent a similar time to the other clusters interacting with each post. These findings can be interpreted in several ways. First, it is possible that in this context of a discussion composed of relatively short messages (192 words on average) the time they spent per message (just under five minutes) was sufficient for deep engagement; thus breadth of engagement is a more important measure of listening quality. Second, it is possible that even learners in this "mastery" cluster are not engaging with posts at a very deep level and interventions to support this process are needed. Finally, it is possible that they engaged deeply with some posts but not others and this behavior was not revealed by the average time. In addition, similar to learners in the other clusters, Broad Listeners scanned 50% of the posts they viewed, suggesting that this may be a useful or necessary strategy to manage participation in a many-message forum. Future research is needed to understand the motivations behind these listening behaviors. Additional work is also needed to examine the reflective talking behavior observed for this cluster: why are these learners repeatedly returning to their own posts, is this supportive of their learning, and how might such behaviors be encouraged in other learners?

Implications for Practice

Because this line of research is still in an early stage, it would be premature to present strong prescriptions for practice. However, the results of this study do suggest several potential ways to support productive listening behaviors that can be explored with design-based research. First, because Superficial Listeners tend to distribute their participation in multiple short sessions, keeping discussions somewhat compact in duration (e.g. one to two weeks) may help to concentrate their involvement and lead to more coherent participation. Second, because the increased listening observed for Broad Listeners was due to a greater number (not length) of interactions with the tool and individual posts, designing conversations that encourage or require learners to engage multiple discrete times may encourage more extensive listening behaviors. One way we have started to explore this is by requiring posts at discrete times during the discussion week (e.g. Monday/Tuesday, Wednesday/Thursday, Friday/Saturday), though there are certainly other strategies that can be used. Finally, the high number of delayed reviews seen for Cluster 3 (Reflective Talkers) suggests that reflective visitation of one's own posts can be explored as a tool for supporting metacognition and synthesis of learning in online discussions.

Conclusion

Past research on online discussions has focused primarily on contributions to online discussions. However, a conceptual model of online discussions as conversations also necessitates attention to learners' listening behaviors. This study showed that listening represents a very large proportion of the time learners spend in online discussions (almost three quarters of their time in this case). Similar to research on lurkers in online communities (Nonnecke & Preece, 2000) this reinforces the prevalence of often invisible forms of participation and highlights the importance of investigating these behaviors using methodologies such as the click-stream approach employed here. In addition, listening is a substantial and important component of online discussions that can potentially be targeted for design (e.g. via scripting, see Fischer et al., 2007). This study identified three distinct styles through which learners interacted with online discussions to listen to their peers. While similar to mastery, task and minimalist approaches found for other contexts (del Valle & Duffy, 2007), this study pinpointed the specific listening (and speaking) behaviors that make up such styles in the context of online discussions. Future work can investigate the stability of these styles across different learners and contexts, examine their relationship to discussion quality and learning, explore techniques for early detection of particular styles and research design-based interventions to support productive interactions in online discussions.

References

Arvaja, M. (2007). Contextual perspective in analysing collaborative knowledge construction of two small groups in web-based discussion. *International Journal of Computer-Supported Collaborative Learning, 2*(2-3), 133-158.

Barab, S. A., Bowdish, B. E., & Lawless, K.A. (1997). Hypermedia navigation: Profiles of hypermedia users. *Educational Technology Research and Development, 45*(3), 23-41.

Barab, S. A., Kling, R., & Gray, J. H. (2004). (Eds.). *Designing for virtual communities in the service of learning*. Cambridge, UK: Cambridge University Press.

Boulos, M. N., & Wheeler, S. (2007). The emerging web 2.0 social software: An enabling suite of sociable technologies in health and health care education. *Health Information and Libraries Journal, 24*, 2-23.

del Valle (2006). Online learning: Learner characteristics and their approaches to managing learning. (Unpublished doctoral dissertation). Indiana University, Bloomington Indiana.

del Valle, R., & Duffy, T. M. (2007). Online learning: Learner characteristics and their approaches to managing learning. *Instructional Science, 37*(2), 129-149.

De Wever, B., Schellens, T., Valcke, M., & Van Keer, H. (2006). Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. *Computers & Education, 46*(1), 6-28.

Fischer, F., Kollar, I., Mandl, H., & Haake, J. (Eds.). (2007). *Scripting computer-supported communication of knowledge. Cognitive, computational, and educational perspectives.* New York, NY: Springer.

Herring, S. (1999). Interactional coherence in CMC. *Journal of Computer-Mediated Communication, 4*(4).

Hewitt, J. (2003). How habitual online practices affect the development of asynchronous discussion threads. *Journal of Educational Computing Research*, *28*(1), 31-45.

Hewitt, J. (2005). Toward an understanding of how threads die in asynchronous computer conferences. *Journal of the Learning Sciences, 14*(4), 567-589.

Hewitt, J., Brett, C., & Peters, V. (2007). Scan rate: A new metric for the analysis of reading behaviors in asynchronous computer conferencing environments. *American Journal of Distance Education, 21*(4), 215-231.

Kear, K. (2001). Following the thread in computer conferences. *Computers and Education, 37*(1), 81-99.

Luppicini, R. (2007). Review of computer mediated communication research for education. *Instructional Science, 35*(2), 141-185.

Nonnecke, B., & Preece, J. (2000). Lurker demographics: Counting the silent. *Proceedings of CHI 2000.* The Hague: ACM Press.

Stahl, G. (2003). Building collaborative knowing: Elements of a social theory of learning. In J. W. Strijbos, P. Kirschner & R. L. Martens (Eds.), *What we know about CSCL in higher education* (pp. 53-85). Amsterdam, NL: Kluwer.

Suthers, D. (2006) Technology affordances for intersubjective meaning making: A research agenda for CSCL. *International Journal of Computer-Supported Collaborative Learning, 1*(3), 315-337.

Thomas, M. J. W. (2002). Learning within incoherent structures: the space of online discussion forums. *Journal of Computer Assisted Learning, 18*, 351-366.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association, 58*(301), 236-244.

Acknowledgments

This work was supported by the Social Sciences and Humanities Council of Canada.

Socially Constructive Interaction for Fostering Conceptual Change

Moegi Saito, Naomi Miyake, Consortium for Renovating Education of the Future, University of Tokyo, Japan
Email: saitomoegi@coref.u-tokyo.ac.jp, nmiyake@p.u-tokyo.ac.jp

Abstract: In order to guide the design and assessment of successful intentional conceptual change at school, we examined third graders' learning processes using a series of lessons taught by a highly experienced teacher. The lesson unit consisted of 11 problems whose answers were to be predicted and discussed one at a time. All 21 students in the class successfully learned the target concept. Our analysis revealed that they started by confirming their rudimentary ideas taken from daily experiences, and then gradually developed theory-like concepts by repeatedly discussing their predictions to be confirmed with experiments. Toward the end of the unit, individual students could express their understanding in their own words, indicating that they successfully integrated their folk knowledge with scientific thinking. Further analysis of the dialogue contents indicated that some small-group discussions were effectively monitored and utilized by the rest of the class, to assure achievement of the entire class.

Introduction

This presentation aims to clarify the structures of successful classroom discussion that support learners in changing their folk knowledge into scientific understanding. It has been suggested that sequential cumulative discussion across students has strong positive effects in helping learners adaptively expand their understanding (e.g., Hatano and Inagaki, 1991; Schwartz and Martin, 2004; Miyake, 2008). However, the details of such classroom discussion have rarely been analyzed fully to explore whether a specific structure leads to adaptive conceptual change. In this paper, we report the results of our analysis of classroom discussions of 12 consecutive lessons of a third-grade class of 21 students, regarding their understanding of the physical identity of objects (the inability of two objects, like air and water, to share identical physical space). The lessons were conducted according to the framework of Hypothesis-Experiment-Instruction (HEI) (Hatano and Inagaki, 1991; Miyake, 2008). Overall, our analysis revealed that the students' discussion developed according to a specifiable structure, which we could formulate as a socially expanded version of two-person constructive interaction (Miyake, 2008). This analysis should help develop better CSCL systems for promoting collaborative learning in daily classroom practices.

Research Context: Hypothesis-Experiment-Instruction

In this section, we explain the framework of HEI, created by Itakura in 1963, and the target content of the "Air and Water" unit (Itakura, 2000) used as our research context. HEI is a strategy to teach basic scientific concepts. An HEI lesson consists of a set of strategic procedures for discussion and a problem. The students predict and discuss the answer to this problem, which is revealed as a result of a relevant experiment at the end of the discussion. An HEI unit consists of multiple problems, or experiments, carefully ordered to guide the development of scientific concepts underlying the problem set. To practice a unit, a teacher uses a problem sheet for each experiment, which explains the experiment and possible answers. Students are expected to integrate the results of the experiments in their own way to formulate individualized hypotheses, or rudimentary scientific concepts owned by individual students.

For a typical HEI problem, students follow the four steps (Itakura, 1997).

1) The teacher demonstrates the procedure of the experiment, and the students write their predictions of its outcome by choosing one of the alternatives. The results of the choices are made explicit when the teacher counts and writes on the blackboard the number of students who chose each alternative.
2) Students openly explain their choices to the class, so that all class members can hear the explanations. During this discussion, the students are allowed to express their ideas at will, including staying silent, and to change their choices if they wish to do so.
3) When this discussion naturally comes to an end, the experiment is conducted either by the teacher or by the students in small groups.
4) After the experiment, students write comments individually.

These four steps are repeated in a unit, in the order of the problems. All the HEI units were developed by Itakura and his core group, and have been tried out and refined by teachers in real classrooms.

The "Air and Water" Unit

The "Air and Water" unit consists of the 11 problems presented in Fig. 1, with explanations in Table 1.

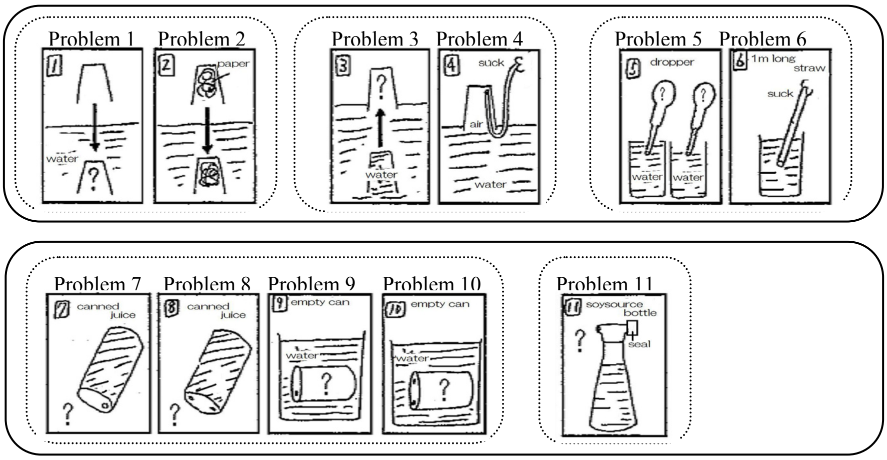


Figure 1. Depictions of the 11 Problems in the "Air and Water" HEI Unit.

Table 1: Wording of the 11 problems in the "Air and Water" HEI unit.

Problem 1. When an empty glass is pushed into water upside down, will the water come into the glass? Problem 2. If you place a crumpled piece of paper in the glass and do the same as in Problem 1, will the paper get wet?
Problem 3. An upside-down glass with water inside is in the water. When you lift it up through the surface of the water, what will happen to the water in the glass?
Problem 4. What will happen when you suck air through a straw from an upside-down glass in the water?
Problem 5. Which dropper sucks more water, one whose tip is deep in the water or one whose tip is shallow?
Problem 6. Can water be sucked through a 1m straw?
Problem 7. A can of juice has just one hole on top. When the can is turned upside down, will some juice come out of it?
Problem 8. Will some juice come out of a can that has two holes on its top and is turned upside down?
Problem 9. Suppose you put the can used in problem 8 deep into the water, keeping your finger tight on one of the holes. Will some water go into the can?
Problem 10: What will happen to the can in problem 9 if you let your finger go?
Problem 11: Will some soy sauce come out of its container if you put your finger onto the hole on its top?

These 11 problems can be classified into two subsets, Problem 1 (P1) to P6, and the rest. The first set includes problems whose answers are justifiable with daily experiences. Their answers are easy to imagine because the problems ask for direct sensing, as "visible" bubbles (P1, 2, 3, and 5) or "sense of mouth" (P4 and 6). In contrast, the situations of P7 to P10 are difficult to imagine because they do not often occur in the students' daily lives. These problems require learners to rely on their newly formed hypotheses, from predicting and observing the experiment results in the previous problems. The last problem, P11, can be answered by relying on either daily experiences or newly learned understanding, or both.

The problems in a unit are ordered so that answering the latter problems encourages active integration of the experiences gained through the preceding problems. The hypotheses required to answer the latter problems require higher-order abstractions of the experiences gained through the former problems. Miyake (2009) proposed a four-stage model of conceptual change (Table 2), where the concepts at both levels 1 and 2 could be formed based on individual experiences, while those at levels 3 and 4 require integration of different perspectives socially distributed among different individuals. According to this model, in the "Air and Water" unit, P1 to P6 can be answered with understanding at levels 1 and 2, while answering problems after P7 requires integration of different perspectives among learners.

Table 2: Four-stage model of conceptual change.

Level 4	Scientific concept, created and shared in the scientific community
Level 3	Confirming explanation by integrating many ideas, including "textbook" scientific concepts
Level 2	Rules of thumb created by accumulating one's own (yet many) perspectives from different situations
Level 1[t]	Explanation or "theory" like folk concept based on one incidence

Hypotheses

Judging from the model in Table 2, we hypothesized that answering problems after P7 required level 3 and up collaborative integration of one's own multiple experiences and those of others. In order to test this hypothesis, we analyzed the predictions made before and after the discussion, accuracy of the predictions, and number of utterances made by the class as well as by each individual student during the discussion and after the experiments, for each of the 11 problems. More specifically, we tested whether the students' predictions of the experimental results for P1 to 6 were different from those for P7 to 11. While the prediction of correct answers tended to be high for P1 to P6 even before the discussion, it could be low for the earlier problems of the latter set (P7 and 8), and gradually increase toward P11. The predictions would also be more discrepant before and after the discussions for the latter set of problems. We also tested whether we could identify collaborative construction of abstracted concepts of level 3 and up in the discussion of the problem set of P7 to P11.

Data Analysis and Results

The Data

The data came from the "Air and Water" unit in an HEI class conducted in May and June of 2002. Twenty-one third graders participated in 12 lessons taught by a highly experienced HEI teacher, Yuko Saito, who voluntarily kept records of the students' answers and discussions using hand-written notes and voice recordings.

Figure 3 indicates the shift of predictions over the 11 problems. Each box corresponds to a problem in the unit, in order. In each box, the white bars indicate the number of choices made before discussion; the choices included the alternative answers of A, B, C, and D; the circled one (e.g., B for P1), is the correct answer. The black bars indicate the number of choices made after the discussion but before the experiment.

Predictions of Answers

For P1 to P4, the students tended to choose the correct alternative from the beginning and to stay with it after discussion (Fig. 3). For P5 and 6, they experienced a slight discrepancy between the two choices, but tended to choose the correct answer after discussion. This difference between P1 to P4 and P5 and 6 may be due to the ease of finding reliable reasons for choices: P1 to 4 could have been answered by remembering ordinary experiences at home (e.g., taking a bath). In P5 and 6, the problem situation shifted away from such familiar scenes to those using tools often found in a science class. With P7, the pattern suddenly changed, with no one choosing the correct answer even after discussion. The same phenomenon happened again for P8; however, with P9, the students recovered and were once again able to predict correctly how the experiment would turn out. They were shaken again on P10; however, they were all able to predict the correct answer to the last problem, P11. The overall pattern indicates that at the end of the unit, all the students in this class somehow successfully formed a rudimentary concept of physical identity, at least with regard to whether or not air and water could share the same physical space. The shift pattern matched what we hypothesized in Section 2.

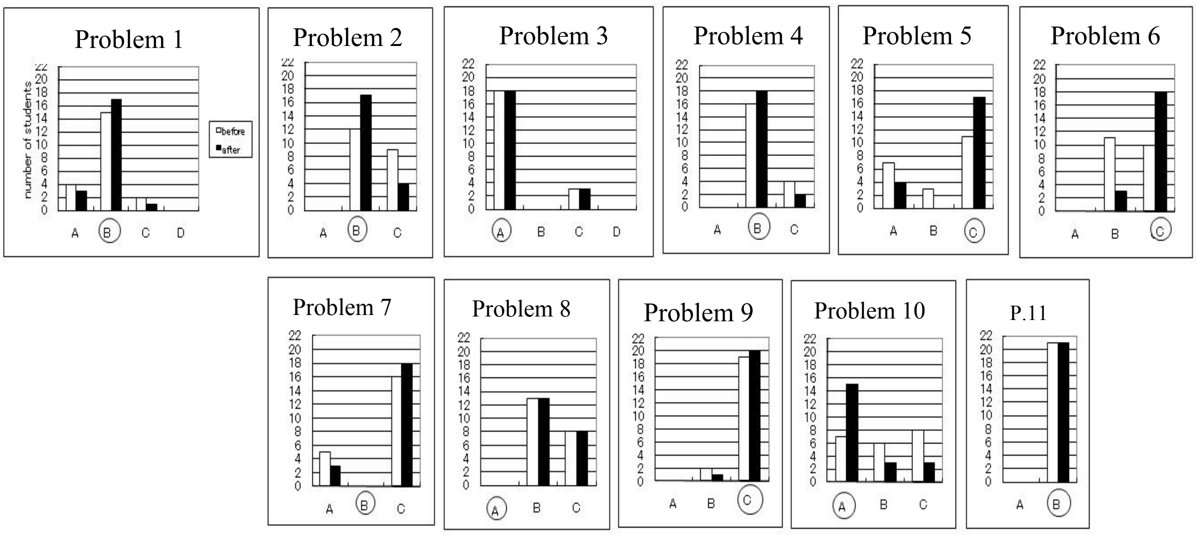


Fig. 2. Student's Choices of Alternative Answers for Each Problem in the "Air and Water" HEI Unit.

Shift Pattern of Predictions

If the shift pattern of predictions matched our hypothesis (1), we could expect to observe a corresponding trend in the number of level 3 and up utterances in our model in Fig. 2. We analyzed the contents of the student discussion to test this hypothesis. For an utterance to be classified as level 3 and up, it had to include expressions about the inability of air and water to share the same physical space (e.g., "The seal stops the air. When the air can't move, the water won't move, either" for P11). An utterance such as "I have tried such an experiment with a wash bowl in the bath" for P1 would not be level 3 or above. The correspondence rate of coding by two coders was 94%.

Figure 3 depicts the pattern we observed. The bars for each problem indicate the number of utterances at all levels. The black bars at the bottom indicate the number of level 3 and up utterances for each problem. As we hypothesized, for P1 to P6 the students did not talk much at level 3 or up, though some could express such understanding as early as P1. For P7, again as we expected, the number of level 3 and up utterances was close to zero, with only two students expressing such understanding. Thereafter, the number rose sharply for P8 but decreased toward the end of the unit, as the total number of utterances also decreased. This trend differed from our expectation, yet the level 3 and up utterances for P7 to 11 totaled 63, which was much higher than the total of 38 for P1 to 6. The average number of such utterances was 6.3 for P1 to P6, and 12.7 for P7 to P11.

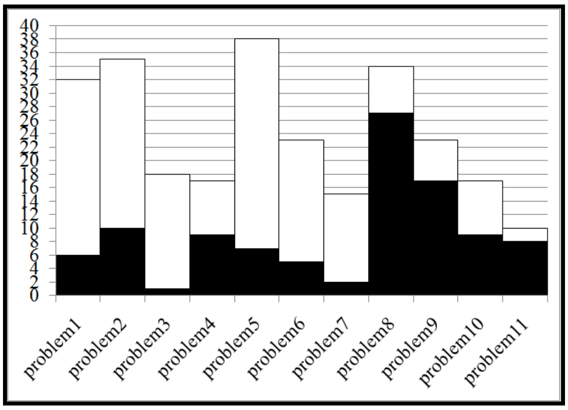


Figure. 3. Number of Students' Utterances during Discussion in the "Air and Water" HEI Unit.

This trend could be interpreted as indicating that most of the students succeeded in changing their folk theory (i.e., rule of thumb based on everyday experiences) into more scientific theory-like understanding by bridging two different types of problems. However, level 3 and up utterances existed even at P7 (though the owners did not make the correct prediction), and at P8 nobody could predict correctly even though many utterances were coded as level 3 and up. Thus, this quantitative whole-class analysis did not seem to provide enough information to confirm our hypotheses. To go one step further, we analyzed each student's utterances.

Coding of Discussion Patterns in Terms of Role Exchange

In order to identify a pattern among the discussions, we used the framework of constructive interaction developed and tested for a two-person joint problem. According to this framework, role exchange of task-doing (uttering ideas) and monitoring (listening) appears to be responsible for abstracting understanding (Shirouzu et al., 2002). It has been proposed that this mechanism happens socially to induce collaborative conceptual change (Miyake, 2008), or to help learners collaboratively change their folk theory into a scientific concept, as indicated in the model described above (Miyake, 2009). In order to test this idea, we coded the patterns of turn-taking during the discussion of the 12 lessons into two categories of role exchange: individual sequence and group dialogue, defined as follows.

Individual sequence Learners take turns to express their ideas in succession. Illustratively, utterances would move as A>R>F>C>H…. This is a linear successive role exchange in which a single person's verbalization is handed down sequentially to the next person. Though there is not much chance for constructive interaction among the speakers, this process assures an equal chance for all class members, and thus is sometimes preferred by teachers (CoREF, 2010).

Group dialogue Role exchange can happen socially, involving groups. In a class discussion, a small group of two to three students may engage in a positive discussion, exchanging roles of task-doing and monitoring among them for a while. Yet at the same time this group as a whole could be assuming the role of **task-doing** to the whole class, possibly working as a core or leading group of the lesson, whose discussion can be **monitored** by the other students in the class. When this group's discussion comes to a halt for some reason, a different group of students takes over the task-doing role by starting their own discussion to be monitored by the rest of the students. Such a shift of speakers exhibits a pattern, such as A>B>A>B followed by C>D>C>D, then by B>D>E>F>D>F.

Which pattern leads the class to a more successful conceptual change? If teachers' preferences are relevant, we would expect to observe more individual sequences in successful lessons such as those we analyze here. A prediction based on the rigid formulation of the constructive interaction framework also favors individual sequences because it requires each participant to experience both roles. Yet we can also expect group dialogues to lead to successful learning, as we observe a small group of selected class members engage in a long heated class discussion, which is monitored rather positively by the other members. If our analysis identifies this group-based pattern, this result may lead to new understanding of how to design and support such activities more positively than before.

Analysis of Role-exchange Patterns in Group Discussion

We analyzed the role-exchange pattern in discussions with the following two procedures.

Procedure 1 We counted the utterances of each individual student who participated in our data, to determine if any bias existed in the frequencies of role exchange. The students were assigned letters A to U based on the number of utterances, where the most frequent talker was A.

Procedure 2 For each of problems P7 to P10, we tabulated the shift to the task-doing role identified as a talker of her/his opinion. We also analyzed utterances relevant to conceptual change.

The data consisted of 263 utterances, 9 of which were uttered by the teacher. Because all her utterances were made in an effort to control the progress of the student discussion (e.g., “We don’t have enough time. Let’s experiment”), they were not coded, leaving 254 student utterances for this analysis.

Role-exchange Frequencies

Figure 4 lists the number of utterances for each student in the class, from the one who talked the most (top) to the one who talked the least (bottom). Students are identified alphabetically. Figure 5 indicates the order of students who talked for each problem. It is possible to identify the role exchange pattern (individual-based or group-based), with demographic information on whose comments were followed by whom, when, and how often. In order to identify the interaction patterns, we looked for sequences in which two to four persons iterated their utterances, and coded them as group dialogues (denoted with black cells in the graph). During these group-based exchanges, we assumed that some constructive interaction occurred among the members, by exchanging roles of task-doing and monitoring. This iterative pattern was identified not only by the orderly exchange of utterances but also by content continuity. Utterance numbers 17 to 26 of P8, skipping M’s utterance at number 18, is such a case. The white cells indicate no apparent group dialogue, thus representing individual sequences.

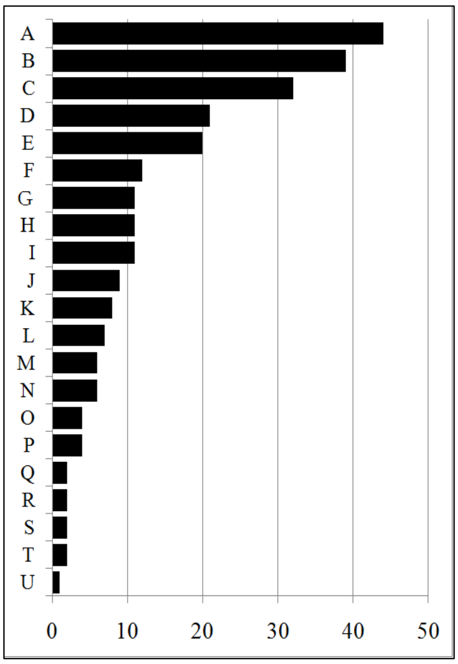


Figure. 4. Number of Utterances for Each Student.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
1	J	N	K	I	I	J	C	B	I	K	F
2	F	D	J	H	K	D	N	K	F	N	B
3	C	?	A	F	D	E	J	F	M	D	A
4	B	M	H	G	C	P	H	L	G	A	E
5	H	H	B	A	H	N	E	S	C	H	G
6	D	F	L	?	N	I	G	P	A	M	M
7	A	B	G	A	A	K	A	M	D	O	A
8	G	D	E	?	L	Q	B	D	C	I	F
9	E	E	A	A	B	C	D	J	A	L	A
10	B	A	L	D	D	K	P	C	C	T	F
11	U	T.U.	A	A	J	E	T	A	A	G	
12	F	B	B	H	E	G	J	E	C	E	
13	A	?	A	B	I	A	E	B	A	B	
14	B	E	E	A	K	O	B	A	C	C	
15	D	O	A	T.U.	O	B	T.U.	E	L	B	
16	B	P	E	all	H	C	A	B	?	A	
17	D	N	A	T.U.	R	T.U.		C	E		
18	B	R	E	S	A	J		M	C		
19	A	?		I	D	?		B	B		
20	B	C		A	B	A		A	A		
21	A	B		C	C	C		C	I		
22	B	D			B	T.U.		B	Q		
23	K	B			?	H		C	C		
24	C	C			A	D		A			
25	B	B			D	A		B			
26	C	C			C	T.U.		C			
27	B	D			A			I			
28	I	A			T.U.			F			
29	B	B			A			H			
30		C			E			E			
31		B			D			C			
32		C			L			B			
33		B			A			F			
34		?			B			A			
35		C			D						
36		F			C						
37		T.U.			G						
38					B						
39					J						
40					E						

Figure. 5. Role-exchange Patterns for Each Problem.

We infer from the data in Fig. 4 that if any group-based task-doing occurred, the chances were high that the group included students A, B, and C, and possibly D and E. The graph in Fig. 5 also indicates that during these 11 collaborative problem-solving lessons everyone talked; thus, every student played a task-doing role at least once.

In Fig. 5, the percentages of black cells in each column are lower for P1 to P6 (44.8% for P1, 35.1% for P2, 33.3% for P3, 23.8% for P4, 25% for P5, and 0% for P6) than for P8 to P11 (44.1%, 39.1%, 0%and 40%). This result may indicate that for this class, it was necessary for the students to accumulate different perspectives, or expressions based on individually different experiences in the former phase, in order to move on to more constructive group-based interaction regarding a focused sharable topic on which selected members exchanged roles, possibly leading them to raise their understanding (Shirouzu et al., 2002). This group dialogue could have worked as group-based ***task-doing***, whose discussion was ***monitored*** by the rest of the class members.

To analyze the distribution patterns of the black cells in Fig. 4, we must rely more on our observations and interpretation of constructive dialogues. With some caution, we made the following conclusions. Because P7 was a new problem for everyone, each student expressed her/his own ideas, resulting in a long sequence of individual sequences. During P8, after watching what happened to the juice in the can with one hole when it was turned upside-down, the same students appeared to be able to focus on selected aspects of the phenomenon, thus starting some group dialogue on such topics. Two such group discussions could have been highly positive, judging from the number of people involved and the length of the discussions. However, this group task did not seem to have worked fully to adaptively deepen the understanding of the members in the group or the rest of the class, because none of them correctly predicted the result of the experiment (Fig. 2). For P9, students A and C again engaged in group dialogue, as if they had picked up on their interaction for P7, providing a task-doing group for the class to monitor. Because most of the class members were able to answer P9 correctly, this group task-doing could have been effectively monitored by the class members; most of them understood the content of the discussion, and possibly integrated it to deepen their understanding. After this socially implemented constructive interaction between the core groups who were doing the task and the other members who were monitoring, each individual student was ready to express her/his own understanding in her/his own expressions, which led to sequential individual sequences (P10 column in Fig. 5).

Content Analysis of P7 to P10 Discussions

We examined the contents of utterances during the student discussions of P7 to P10, to confirm whether the contents would fit our interpretation of the Fig. 5 patterns.

Table 3: Excerpts of typical utterances during the discussion session of P7.

3	J	I want the juice to come out fast.
…	…	…
7	A	The can has only one hole so the air can’t escape.
8	B	When I punched a small hole in a milk carton, the milk dripped out continuously.
…	…	…

(N.B. Numbers in the left-most column are orders of the utterances in this session. The letters in the second-to-the-left column denote students, as in Fig. 4.)

In this phase, as indicated in Table 3, students expressed their ideas independently, literally in the fashion of “one after the other.” The third and eighth utterances appear to have been based on the speakers’ daily experiences.

Table 4: Excerpts of group-based dialogues during the discussion session of P8.

1st Group-based dialogue

11	A	The juice in the can would block the air from coming in.
12	E	I think the air would enter from both holes and block the juice.
13	B	The air would enter from both holes and pass through the juice in the can.
14	A	Oh yes! The air will turn into bubbles and pass through the juice!
15	E	If the air blocks both holes, where can the juice come out from?
16	B	When the air enters into the can, the juice would come out. But I don’t know how.

2nd Group-based dialogue

17	C	The air won’t enter the can because it is filled with juice.
18	M	Oh! I got it!
19	B	C didn’t listen to me. I said, “The air passes through the juice in the can.”
20	A	The can has no space. It is filled with juice.

21	C	The can is filled with juice. So the air can't break into the can. Why do you think the air can enter the can?
22	B	The can has some space on top.
23	C	I don't think cans of juice have any space.
24	A	Yes, I think so, too.
25	B	Even a little space is enough for the air to enter.
26	C	The can is filled with juice. So the air will never break into the can!

In both of these group dialogues, the group members clearly shared a focus of discussion, taken not directly from their daily experiences but from their observation and interpretation of previous experiments(see Table4). Furthermore, we observed that while the first discussion tended to be about general movements of the air and the juice, the topic of the second discussion was more focused, with more vividly imagined details of the possible phenomena. Judging from this shift of content, we assumed that the students engaged in the second discussion already shared a basic understanding of the target scientific concept that air and water cannot share the same physical space. We also assumed that this discussion was monitored by the other class members, and that each of them compared and integrated the discussion expressions and their own thoughts. During this phase, the listeners monitored the discussion, while students A, B, C, and E expressed and assessed their own understanding.

Table 5: Excerpts of a group-based dialogue during the discussion session of P9.

5	C	Half of the air would escape from the can when it lies sideways.
6	A	I think that air tends to rise. So air won't escape from the hole punched on the lower side of the can.
7	D	What do you mean?
8	C	The can will be put deeply into the water. So I think we don't have to worry about the position of the hole.
9	A	The base of the can would push back the water.
10	C	When we tipped the glass in <Problem 2>, I remember, the glass was filled halfway with water.
11	A	The can with a hole doesn't have a wide mouth like the glass.
12	C	Small bubbles can pass through even a small hole of the can.
13	A	Anyway, I think the position of the can has nothing to do with the amount of water.
14	C	I think that all the air would escape from the can when it stands up. So half of the air would escape from the can when it lies sideways.

Again the students discussed concrete aspects of the problem situation (e.g., whether the can would be standing up or lying sideways, and the position of the hole)(see Table5). Such details were not explicitly expressed in the problem statements; therefore, the students created detailed, concrete images to clarify how much they understood. This activity also appears to help both the students who were discussing as task-doers and those who were listening as monitors, as they went back and forth between scientific abstract expressions (e.g., "air escapes" and "water fills") and concrete images (e.g., of the can, the juice, and the water), as if they were tying their scientific understanding to their daily life experiences.

Table 6: Excerpts of typical utterances during the discussion session of P10.

1	K	The can has two holes. So the water will replace the air that filled the can.
4	A	The air will escape from one hole, and the water will come in from the other hole.
14	J	The air will powerfully block the movement of the water.
17	B	The air will escape from the two holes. After the movement of the air, the water will come in from the two holes.

For P10, the students chose to express their understanding individually(see Table6). Not only students A and B, who served as core task-doers by phrasing and rephrasing their ideas, but also students K and J, who were relatively silent, verbally predicted (K) or made some causal-like expressions (e.g., student J's implication that air has power). Group-based dialogues that occurred in P8 and 9 discussions were monitored by these students, and possibly helped them integrate and abstract their own and others' ideas; as a result, their ideas came closer to scientific concepts.

Here we summarize our interpretation of the socially constructive interaction we observed in the protocol. During the P7 discussion, as we inferred from the role-exchange pattern in Fig. 5, students expressed their individual observations, because the problem situation was new and they did not share a common focus. When they moved on to P8, after exchanging different ideas expressed in various ways and observing the correct answer to the previous problem, some students (e.g., A) were able to express their rules of thumb rather abstractedly, in a top-down fashion (e.g., "Water cannot come out because there is air outside."). When such a statement was questioned by another student, a group dialogue or possibly constructive interaction between the two occurred; the rest of the class monitored this interaction as group task-doing. As the situation progressed, another student, who thought through concrete images of the juice in the can and of the water surrounding it, challenged the abstracted expression and asked for more concrete images. The result was more group task-doing, which was also monitored by the class. This monitorable constructive interaction between A and C, between the abstraction and the concrete images of the set of air-water exchange phenomena, helped each class member integrate the discussion with her/his own ideas toward the end of P9, leading them all to correct answers. For P10, all the students were able to express their individual understanding or concept in their own styles, rather sequentially. This pattern was the same as for P6 and P7, but the levels of the content according to the conceptual change model in Table 2 were different. The students' utterances contained more level 3 expressions during P10 discussion than during P6 and 7 discussions.

Discussion

In the discussion protocols of 21 third graders going through 11 carefully ordered HEI problems in an "Air and Water" unit, we observed successful socially constructive interaction among both individuals and small groups. We also found that the role exchange of task-doing as speakers and monitoring as listeners supported the entire class to raise each individual's understanding of the scientific concept, the target of the learning, as predicted from two-person constructive interaction studies. The 21 students in the class, in a facultative atmosphere where everybody felt free to talk, expressed diverse individual ideas on encountering new problems. The same students could also argue on a topic of focus after sharing some common experiences. When they argued about a selected topic, a small group of two to four members engaged in constructive interaction to form a task-doing group whose discussion was monitored by the class and sometimes integrated by some monitoring members. This group-based constructive interaction occurred several times, with different core members engaging in the task that was monitored by the rest of the class. The class exhibited high performance with everybody correctly answering the last question at the end of this highly social constructive interaction. This result indicated that such complex collaboration is an effective and enjoyable practice that promotes scientific conceptual change.

We plan to analyze similar sets of class discussion data taken from other units of HEI classes and other collaborative classes, in order to understand the conditions for successful discussion, in an effort to design better technological support for the promotion of intentional conceptual change, as well as collaborative learning skills.

References

CoREF , (2010) *A summary report of CoREF at University of Tokyo, Year 2010.* CoREF, December, 2010.

Hatano, G. & Inagaki, K. (1991). Sharing cognition through collective comprehension activity. In Resnick, B. Levine, J. M. &Teasley, S. D. (Eds.). *Perspectives on socially shared cognition*. 331-348. APA

Itakura, K. (1997). Kasetsu-Jikken-Jugyo no ABC, Dai 4 han. (*The ABC of the Hypothesis-Experiment-Instruction: Invitation to enjoyable classes, Ver.4.*) Tokyo: Kasetsu-Sha. [in Japanese]

Itakura, K. (2000). *The workbook of Hypothesis-Experiment instruction <Air and Water>*.Tokyo:Kasetsu-Sha.

Miyake, N. (2008). Conceptual change through collaboration. In S. Vosniadou, (Ed.) *International Handbook of Research on Conceptual Change.* 453-478. New York: Routledge.

Miyake , N. (2009). *Conceptual change through collaboration, Paper presented at AERA 2009,* San Diego.

Schwartz, D.M. & Martin, T. (2004). Inventing to prepare for learning: The hidden efficiency of original student production in statistics instruction. *Cognition & Instruction, 22*, 129-184.

Shirouzu, H. & Miyake, N.,& Masukawa, H. (2002). Cognitively active externalization for situated reflection. *Cognitive Science*, 26, 469-501.

Effects of Using Multiple Forms of Support to Enhance Students' Collaboration during Concept Mapping

Anushree Bopardikar*, Dana Gnesdilow*, Sadhana Puntambekar, University of Wisconsin – Madison, 1025 West Johnson Street, Madison, WI 53706,
Email: bopardikar@wisc.edu, gnesdilow@wisc.edu, puntambekar@education.wisc.edu

Abstract: Collaborative learning provides opportunities to facilitate students' knowledge construction. However, students may face challenges while constructing meaning through dialogue. This study explored multiple forms of support to foster collaboration that would lead to a deeper understanding of science ideas and high-quality group concept maps. Ninety-five students from four 8th grade science classrooms participated in this study. Using a 2 x 2 experimental design, we examined the effect of teacher modeling of prompts and the effect of meaningful individual preparation prior to collaboration on the quality of students' concept maps. Statistical analyses revealed that groups that prepared individual concept maps and received teacher modeling of prompts generated the highest quality concept maps. Further, teacher modeling of prompts was only found to be effective when combined with the individual preparation of concept maps prior to collaboration. Implications for classroom practice and future research are discussed.

Research from collaborative learning in classrooms and Computer Supported Collaborative Learning settings (CSCL) has found that engaging students in collaboration provides opportunities to facilitate their knowledge construction. Students involved in collaboration may learn through generating elaborate questions, explanations, and arguments (e.g., Choi, Land, & Turgeon, 2005; Weinberger, Stegmann, & Fischer, 2006). However, other research reveals that constructing meaning through dialogue often involves challenges. Students may not always interact effectively and may have difficulty extending one another's ideas and sustaining discussions (So, 2007). Our own prior research has found that students do not always engage in elaborate science conversations in groups (Gnesdilow, Bopardikar, Sullivan, & Puntambekar, 2010). Mercer (2000) has explained that people use their past social and cultural experiences to help them to understand the ways that particular dialogic interactions unfold. Students may not be able to draw from such resources during conversations with each other if their prior school experiences have provided them with little time to talk with peers and has focused more on didactic forms of instruction and discourse.

This paper focuses on collaborative concept mapping as a specific context for researching and enhancing students' collaboration. Concept maps are graphic organizers that represent knowledge by showing relationships between concepts. A concept map includes at least two concepts connected by a linking word which explains the relationships between the concepts (Novak & Cañas, 2008). A collaborative concept map is created when students collectively generate a concept map that may reflect the understanding of the group. Collaborative concept maps may promote social thinking and metacognition among students (Roth & Roychoudhury, 1992), and engage students in elaborate science discussions to co-construct scientific meanings (van Boxtel, van der Linden, Roelofs, & Erkens, 2002; White & Gunstone, as cited in Edmondson, 2000).

To facilitate collaborative concept mapping and address the challenges involved in collaborative knowledge construction, there is a need for supporting students' collaboration. Classroom studies on collaborative learning have provided students with prompts for generating high-level science questions and explanations (Coleman, 1998; King, 1999), and instruction and training to students in various group skills, prior to collaboration (Mercer, 2000). In CSCL settings, studies have utilized software-based tools to help students construct elaborate science explanations (Vattam, Kramer, Kim, & Kolodner, 2007), provided online metacognitive support to promote students' critical discussions (Choi et al., 2005), and scripts to enhance students' online discussions (Weinberger et al., 2006). Furthermore, in classroom and CSCL settings, studies have stressed the important role of the teacher in moderating students' discussions and helping students develop collaboration skills (Mercer, 2000; Schwarz, Asterhan, & Gil, 2009).

In this study, we utilized *multiple* forms of support to facilitate students' knowledge construction during collaborative concept mapping. We examined how two supports, (i) meaningful individual preparation and (ii) teacher modeling of prompts prior to collaboration, affect the quality of students' collaborative concept maps. We will explain the forms of support in more detail in the methods section.

Methods

Participants and Instructional Context

We conducted this study in four eighth grade science classes taught by the same teacher in a public school near a mid-sized US Midwestern city. Twenty-five groups, with a total of 95 students, participated in this study. These students had worked in collaborative groups in their science class for about three months prior to this study. Students then engaged in a seven week design-based science curriculum, called CoMPASS, to learn about work and energy in simple machines through completing a set of five design challenges (Puntambekar, Stylianou, & Goldstein, 2007). Students collaborated in the same small groups of three or four during the CoMPASS unit to generate predictions and questions, to explore information on a hypertext system, and to perform physical and computer-based experiments. Students also constructed several individual and collaborative concept maps during this unit, which were used for both instructional and assessment purposes.

Study Treatment and Comparison Groups

Our study focused on supporting collaborative knowledge building during group concept mapping by investigating the effects of *multiple* forms of support for students' collaboration. Based on previous research, we designed two forms of support to enable students to engage in collaborative knowledge building through concept mapping. The first form of support was teacher modeling of prompts to involve students in more elaborate science discourse. This was grounded in sociocultural approaches to learning which emphasize cultural tools and practices, such as discourse, during collaborative activities (Vygotsky, 1978). Consistent with studies on the teacher's role in facilitating students' collaboration, we included teacher modeling of these prompts for a constructive student dialogue (Mercer, 2000; Schwarz et al., 2009). This support aimed to facilitate students' *discussion* of science concepts when working in peer groups and to provide them with social and cultural experiences prior to collaboration as models of how elaborate science conversations may unfold through the teacher's expert enactment of the prompts (Mercer, 2000). As in prior research by Coleman (1998) and King (1999), our prompts reminded students to question, elaborate, and justify their ideas during dialogue with their peers. We provided students with written prompts and the teacher demonstrated constructive science discussions using these prompts. Specifically, the teacher read each prompt to the class and explained its meaning in her own words. Then she modeled a scenario to elicit students' thinking about how to use the prompts to help their peers elaborate and improve their science ideas to make deeper connections on the map. The prompts were designed to support students on both individual and collaborative levels. One of the individual level prompts was, "justify your thinking about why the concepts or relationships between concepts you contribute are important," and one of the group level prompts was, "ask other group members to justify, give evidence, or support their ideas."

The second form of support explored an individual preparation activity related to the science *content* prior to the collaborative concept mapping activity. This support was intended to help students generate and reflect on the scientific concepts to productively contribute to their collaborative activity. This is similar to the individual use of the Design Diaries to help students plan their ideas before meeting in their groups, so that they could engage more productively during their collaborations (Puntambekar & Kolodner, 2005). We used individual concept mapping as the meaningful preparation activity prior to collaboration. This choice was based on the work of Novak and Cañas (2008) who explain that concept mapping provides a *scaffold* or *template* to help students organize and structure their knowledge. They further associate concept mapping with a constructive process of learning or knowledge creation, which they describe as being different from more rote forms of learning. This choice was also based on a study by van Boxtel, van der Linden, and Kanselaar (2000), who asked students to individually prepare either a concept map or a poster for five minutes before engaging in a collaborative activity. Their findings, though not statistically significant possibly due to the brief time period allotted, showed greater conceptual understanding for students who had time for individual preparation. In our study, we offered students greater time for individual preparation to examine the effect of the nature of individual preparation on the quality of students' collaborative concept maps. We included an individual lists condition to control for time on task between the different conditions. Students had approximately 12 minutes to individually prepare for their collaboration using any classroom resource. We believe that constructing a concept map is a meaningful activity because it requires students to consider different concepts and to represent the relations between these concepts, helping them reflect on and thus be metacognitive about their prior ideas. This metacognitive engagement may help students contribute their ideas towards an elaborate science discourse, which is consistent with research highlighting the role of metacognitive activities in science classrooms (Gunstone, 1994; Hennessey, 2003).

The study involved a 2 x 2 experimental design to examine the effects of teacher modeling of prompts and meaningful individual preparation on the quality of group concept maps. The four conditions were: i) individual map + teacher modeling of prompts; ii) individual map only; iii) individual list + teacher modeling of prompts; and iv) individual list only (see Table 1). We randomly assigned the four classes to the four conditions.

All students in all conditions received the same directions. The teacher emphasized deep connections between concepts and offered suggestions for focal concepts and linking words. Students could use resources from previous class activities for their collaborative concept map. Students worked face-to-face in their small groups for 20 minutes to construct these paper and pencil group concept maps.

Table 1: 2 X 2 study design of teacher modeling of prompts and type of individual preparation activity.

Factors	Teacher Modeling of Prompts	No Teacher Modeling of Prompts
Individual Map	Individual Map + Teacher Modeling of Prompts (6 groups)	Individual Map Only (5 groups)
List	Individual List + Teacher Modeling of Prompts (6 groups)	Individual List Only (8 groups)

To investigate the two forms of support across the four conditions discussed above, we formulated three research questions: 1) Do groups that receive teacher modeling of prompts generate more deep level, elaborate connections, on their group concept maps than groups that do not receive teacher modeling of prompts? 2) Do groups that engage in individual map preparation generate more deep level, elaborate connections on their group concept maps than groups that engage in individual list preparation? 3) Do groups that engage in individual concept maps and receive teacher modeling of prompts generate more deep level, elaborate connections on their group concept maps than groups in all other conditions?

For our first question, we hypothesized that groups that received teacher modeling of prompts would generate more deep level, elaborate connections on their group concept maps than groups that did not receive teacher modeling of prompts. For question two, we hypothesized that groups that engaged in individual concept map preparation would generate more deep level, elaborate connections on their group concept maps than groups that engaged in individual list preparation. Finally, for our third question, we hypothesized that the individual concept maps + teacher modeling of prompts groups would generate more deep level, elaborate connections on their group concept maps than groups in all other conditions.

Data Sources and Analysis

We analyzed 25 collaborative concept maps across the four conditions that students drew at the end of the work and energy unit. We focused on the quality of the propositions, or connections, on the maps. A proposition is two concepts connected by a linking word to form a semantic unit. Each proposition was scored based on a three level coding scheme modified from prior work (Gnesdilow et al., 2010; Puntambekar et al., 2007) (see Table 2). A Level 2 proposition score shows a more elaborate understanding of science concepts than proposition scores assigned a Level 1. The first two authors independently coded 15% of all maps with about a 92% inter-rater reliability. The remaining maps were then scored.

Table 2: Concept maps scoring rubric.

Level	Description	Example
0	Incorrect, ambiguous, or no connection	levers are inclined planes
1	Simple fact, example, definitions, formula, overgeneralization, unelaborated connections, (e.g. increases, decreases, reduces), everyday language acceptable	third class lever is one of three kinds of lever
2	Scientific language, elaborate explanations, specifying conditions for increase or decrease	levers increase MA when the fulcrum is closer to the load than applied force

Results

We examined the effects of the four conditions on the quality of students' collaborative concept maps as follows. First, we performed a Kruskal Wallis test to identify if there were pre-existing differences among the four classes in students' content knowledge. This preliminary analysis was important to establish equivalence among the conditions. Next, we performed tests of equality of proportions to compare the quality of connections on the concept maps in the four conditions by examining the proportion of Level 2 connections on the maps.

Kruskal Wallis Analysis

We performed a Kruskal Wallis test to determine if the four classes (conditions) differed in their content knowledge prior to the final collaborative concept mapping activity. All students individually took an online test of physics knowledge contextualized in the pulley unit. This test consisted of 19 multiple choice and 11 open-ended items. The total score on this test ranged from zero to 45 points. We chose the Kruskal Wallis test for this preliminary analysis because of the small sample size of less than 30 students in each condition and because the assumption of normality was not met in each condition. The test, which was corrected for tied ranks, was not significant, χ^2 (3, N = 95) = 1.65, p = .6481, α = .01 (see Table 3). This result indicated that there was no statistically significant difference in students' pre-existing content knowledge among the four conditions.

Table 3: Kruskal Wallis Test case summary table.

Condition	*n*	Mean Rank
Individual Map & Teacher Modeling of Prompts	27	52.31
Individual Map only	15	65.47
Lists & Teacher Modeling of Prompts	25	53.10
List only	28	55.77

Tests of Equality of Proportions

Main Effects of Teacher Modeling of Prompts vs. No Teacher Modeling of Prompts and Individual Preparation of Concept Maps vs. Lists

We collapsed the four different conditions into the two factors as per our 2 x 2 design to test the effect of each of the two factors. We performed two, two-tailed planned pairwise comparisons for the two factors to test the equality of proportions of Level 2 connections versus Level 1 connections. We focused on comparing the different conditions according to the frequency of Level 2 connections because we had designed the two forms of support to encourage deep and elaborated science connections on students' concept maps. For this study, we did not examine the frequency of connections assigned a score of zero because a score of zero could indicate multiple issues such as misconceptions, ambiguous or no connections, thereby making comparisons between groups problematic to interpret.

We created two, two-way contingency tables to examine the proportion of Level 2 connections by condition (see Tables 4 & 5). The sample size for each of these comparisons was the total number of correct science connections from all concept maps in the study (N = 491). The null hypothesis for these two comparisons was that the groups in the two conditions within each factor would not differ in the proportions of Level 1 and Level 2 connections on their concept maps. The alternate hypothesis was that the groups in the two conditions within each factor would differ in their proportions of these connections.

The first comparison examined the proportion of Level 1 versus Level 2 connections within the first main factor, teacher modeling of prompts vs. no teacher modeling of prompts. The second comparison examined the proportion of Level 1 versus Level 2 connections within the second main factor, individual preparation activity of maps vs. lists. We used the Dunn method to control the Type I error rate at .01 across the two comparisons (reject if $Z \leq -2.81$ or ≥ 2.81). Both comparisons were statistically significant and confirmed our first two hypotheses. For the first factor, the groups that received teacher modeling of prompts generated a higher proportion of Level 2 connections on their concept maps than the groups that did not ($Z = 3.12$, $p < .01$). For the individual preparation factor, the groups that prepared individual concept maps prior to the collaborative activity generated a higher proportion of Level 2 connections on concept maps than the groups that prepared individual lists ($Z = 5.70$, $p < .01$).

Table 4: Contingency table displaying frequency and proportion of level 2 connections on maps within teacher modeling of prompts factor.

Level of Connection	Modeling Of Prompts	No Modeling of Prompts	Total
Level 1	160 (.81)	268 (.92)	428
Level 2	38 (.19)	25 (.09)	63
Total	198	293	491

Table 5: Contingency table displaying frequency and proportion of level 2 connections on maps within the individual preparation factor.

Level of Connection	Lists	Maps	Total
Level 1	244 (.94)	184 (.79)	428
Level 2	15 (.06)	48 (.21)	63
Total	259	232	491

Effects of the Combination of Teacher Modeling of Prompts and Individual Preparation Activity

Because we found significant differences with our main factors, we decided to delve in further to uncover which conditions may have supported students better in creating high-quality group maps. We set up six, two-tailed pairwise comparisons to study each combination of the two factors and compare the proportions of Level 1 versus Level 2 connections among the four conditions (see Table 6). The sample size for each of these comparisons was the total number of correct science connections from all concept maps in the pairwise comparison groups. The first analysis compared the individual map + teacher modeling of prompts condition to the individual list only condition (N = 291). The second analysis compared the individual map + teacher modeling of prompts condition to individual list + teacher modeling of prompts condition (N = 198). The third analysis compared the individual map + teacher modeling of prompts condition to the individual map only condition (N = 232). The fourth analysis compared the individual list only condition to the individual list + teacher modeling of prompts condition (N = 259). The fifth analysis compared the individual list only to the individual map only conditions (N = 293). The sixth analysis compared the individual list + teacher modeling of prompts condition to the individual map only condition (N = 200).

Table 6: Summary contingency table displaying the frequency and proportion of level 1 and level 2 connections on group maps by condition.

Sophistication of Connection	Individual Map + Modeling of Prompts	Individual Lists Only	Individual Lists + Modeling of Prompts	Individual Maps Only	Total
Level 1	81 (.70)	165 (.94)	79 (.95)	103 (.88)	428 (.87)
Level 2	34 (.30)	11 (.06)	4 (.05)	14 (.12)	63 (.13)
Total	115	176	83	117	491

The null hypothesis for all six comparisons was that there would be no difference in the proportion of Level 1 and Level 2 connections on the maps between the conditions. The alternate hypothesis was that groups in the different conditions would differ in their proportions of these connections. We used the Dunn method to control the Type I error rate at .01 across all six comparisons (reject if $Z \leq -3.15$ or ≥ 3.15). The first three pairwise comparisons were statistically significant. The groups in the individual map preparation + teacher modeling of prompts condition generated a higher proportion of Level 2 connections as compared to the groups in the individual list only condition ($Z = 5.02$, $p < .01$), individual list preparation + teacher modeling of prompts condition ($Z = 5.1$, $p < .01$), and the individual map only condition ($Z = 3.38$, $p < .001$). The remaining three comparisons were not statistically significant (see Table 7). These results supported our third hypothesis that the groups in the individual map preparation + teacher modeling of prompts condition would generate the highest proportion of Level 2 connections.

Table 7: Results for six pairwise comparisons testing proportion of level 2 connections between different conditions.

Comparison	Z Score	p-value
Individual Map + Modeling of Prompts vs. Individual Lists Only	5.02	< .01
Individual Map + Modeling of Prompts vs. Individual Lists + Modeling of Prompts	5.1	< .01
Individual Map + Modeling of Prompts vs. Individual Maps Only	3.38	< .01
Individual Lists Only vs. Individual Lists + Modeling of Prompts	.504	> .01
Individual Lists Only vs. Individual Maps Only	-1.62	> .01

Individual Lists + Modeling of Prompts vs. Individual Maps Only	-1.89	> .01

$$Z = \frac{P_1 - P_2}{\sqrt{\frac{P_1(1-P_1)}{N_1} + \frac{P_2(1-P_2)}{N_2}}}$$

P = Proportion of Level 2 Response

Discussion

While research on collaborative learning has shown that student collaboration can provide opportunities to facilitate students' knowledge construction (e.g., Choi et al., 2005; Weinberger et al., 2006), some studies show that students often struggle to construct meaning through discourse and may not always collaborate effectively (Gnesdilow et al., 2010; So, 2007). In attempting to address this issue, we investigated two forms of support to facilitate students' knowledge building during collaborative concept mapping in a science classroom. We provided students with an individual preparation activity to help them reflect on the science content they could contribute to their collaborative activity and teacher modeling of prompts to help them engage in elaborate science discussions. We found that the groups that engaged in the meaningful initial preparation of making individual concept maps generated more elaborated connections on their collaborative maps than groups that prepared individual lists of concepts. We also found that the groups that received teacher modeling of prompts generated higher quality concept maps than the groups that did not. While we found significant main effects for these two factors, a closer examination of the pairwise comparisons revealed that prompts with teacher modeling and individual preparation of concepts maps were significantly more effective only when they co-occurred. The groups that both engaged in individual preparation of concept maps and received teacher modeling of prompts generated the greatest number of elaborated connections on their collaborative maps than all other groups. Thus, our findings show that it was the *combination* of the two forms of support that was the most effective in helping students generate deep and elaborated connections on their collaborative concept maps.

In the absence of an analysis of students' discourse, we can only offer tentative explanations about the underlying reasons for why the combination of individual concept map preparation and teacher modeling of prompts was most effective. Previous research in classroom collaboration and CSCL has identified the important role of the teacher in facilitating students' discourse (Mercer, 2000; Schwarz et al., 2009). We designed the teacher modeling of prompts from a socio-cultural perspective to provide students with models of effective science conversations as social and cultural tools and practices. Whereas one might expect that teacher modeling of prompts would be beneficial under any condition, this was not the case in this study. When teacher modeling of prompts was combined with lists as an individual preparation, we found no statistically significant difference from the conditions where there was no modeling. Instead there seemed to be a synergistic effect when the teacher modeling of prompts and individual preparation of concept maps were combined. We surmise that having students engage in a meaningful individual preparation of content prior to their collaboration may promote students' metacognitive processes, such as reflecting on, organizing, and making connections between their science ideas, and may help them consider the ideas they could contribute to their collaboration. Additionally, the individual concept mapping may have *primed* the students to take full advantage of the prompts because the creation of the maps may have provided students with a *scaffold* or *template* to organize and structure their content knowledge (Novak & Cañas, 2008). Consequently, the students who were given time to individually prepare concept maps were in a better position to take advantage of the teacher modeling of prompts, perhaps enabling them to better articulate these previously reflected upon ideas. To test our tentative explanations, we plan to analyze students' dialogue during collaborative concept mapping.

Our investigation of students' collaborative concept mapping in a classroom follows from previous research in collaborative learning. Studies have addressed challenges in students' collaboration by providing students with prompts to generate explanations and questions, and with instruction and training in several collaboration skills (Coleman, 1998; King, 1999; Mercer, 2000) and software-based tools to help students construct elaborate explanations, online metacognitive support through tips and examples, and scripts to facilitate students' collaboration (Choi et al., 2005; Vattam et al., 2007; Weinberger et al., 2006). Our study shows that *multiple* forms of support for both the content and models for productive collaboration may be necessary to more effectively guide students' dialogue and to help them reflect on their own ideas prior to their dialogue.

Consistent with the theme of this conference, our findings can also be used to inform teachers' practice. The emphasis on multiple forms of support may be pertinent for younger middle school students in complex collaborative settings, where facilitating both the content and the nature of their dialogue may be critical to enhance their learning. Furthermore, the multiple supports we have designed can be integrated in various curricula, are amenable to the teacher's implementation in a relatively straightforward manner, and can also be easily adapted for CSCL and online settings.

Conclusions and Future Research

We conducted this study with four classrooms of a single teacher and our findings represent a modest first step in exploring ways to support students' collaborative learning and concept mapping. Whereas our quantitative analyses helped us explore differences in the quality of concept maps among the four conditions, we plan to conduct a detailed qualitative analysis of students' dialogue in the different conditions to understand how the multiple forms of support influenced students' collaborative interactions, and how these interactions were related to the quality of the concept maps. Future research could also investigate the effects of individual preparation of concept maps before different types of collaborative activities and explore other forms of meaningful individual preparation prior to collaboration and conditions for which giving prompts with teacher modeling may be most beneficial. While we investigated collaborative concept mapping, future research could study these supportive measures for a variety of collaborative activities. Finally, our study showed the benefits of meaningful individual preparation and teacher modeling of prompts for a single collaborative learning session. Future research could examine these supportive measures over sustained periods of time.

References

Choi, I., Land, S. & Turgeon, J. (2005). Scaffolding peer-questioning strategies to facilitate metacognition during online small group discussion. *Instructional Science*, 33, 483–511.

Coleman, E. B. (1998). Using explanatory knowledge during collaborative problem solving in science. *The Journal of the Learning Sciences, 7*(3/4), 387-427.

Edmondson, K. M. (2000). Assessing science understanding through concept maps. In J.J. Mintzes, J.H. Wandersee, & J.D. Novak (Eds.) *Assessing science understanding: a human constructivist view* (pp. 19-40). San Diego, CA: Academic Press.

Gnesdilow, D., Bopardikar, A., Sullivan, S. A., & Puntambekar, S. (2010). Exploring convergence of science ideas through collaborative concept mapping. In K. Gomez, L. Lyons, & J. Radinsky (Eds.), *Learning in the Disciplines: Proceedings of the 9th International Conference of the Learning Sciences (ICLS 2010) - Volume 1, Full Papers, 698-705.* International Society of the Learning Sciences: Chicago, IL.

Gunstone, R. F. (1994). The Importance of Specific Science Content in the Enhancement of Metacognition. In P. Fensham, R. Gunstone, & R. White (Eds.), *The Content of Science: A Constructivist Approach to its Teaching and Learning* (pp. 131-146). Washington, D.C.: The Falmer Press.

Hennessey, M. G. (2003). Metacognitive aspects of students' reflective discourse: Implications for intentional conceptual change teaching and learning. In G. M. Sinatra & P. R. Pintrich (Eds.), *Intentional Conceptual Change* (pp. 103-132). Mahwah, NJ: Lawrence Erlbaum Associates.

King, A. (1999). Discourse patterns for mediating peer learning. In A. M. O'Donnell & A. King (Eds.), *Cognitive perspectives on peer learning* (pp. 87-115). Mahwah, NJ: Erlbaum.

Mercer, N. (2000). *Words & minds*: *How we use language to think together*. London: Routledge.

Novak, J. D. & Cañas, A. J. (2008) The theory underlying concept maps and how to construct and use them, Technical Report IHMC CmapTools 2006-01 Rev 01-2008, Florida Institute for Human and Machine Cognition, 2008, available at: http://cmap.ihmc.us/Publications/ResearchPapers/TheoryUnderlyingConceptMaps.pdf

Puntambekar, S., & Kolodner, J. L. (2005). Toward implementing distributed scaffolding: Helping students learn science from design. *Journal of Research in Science Teaching*, 42, 185–217.

Puntambekar, S., Stylianou, A., & Goldstein, J. (2007). Comparing classroom enactments of an inquiry curriculum: Lessons learned from two teachers. *The Journal of the Learning Sciences*, 16(1), 81- 130.

Roth, W. M., & Roychoudhury, A. (1992). The social construction of scientific concepts or the concept map as conscription device and tool for social thinking in high school science. *Science Education*, 76(5), p 531-557.

Schwarz, B. B., Asterhan, C. S. C., & Gil, J. (2009). Human guidance of synchronous e-discussions: The effects of different moderation scripts on peer argumentation. In C. O'Malley, D. Suthers, P. Reimann, & A. Dimitracopoulou (Eds.), *Computer-Supported Collaborative Learning Practices: CSCL2009 Conference Proceedings* (pp. 497–506).

So, H. (2007). Improving young learners' scientific understanding in CSCL environments. In C. Chinn, G. Erkens, & S. Puntambekar (Eds.), Mice, Minds, and Society. Proceedings of the Eighth International Computer Supported Collaborative Learning Conference (CSCL 2007): Vol 8 (2) (pp. 645-647). International Society of the Learning Sciences, Inc.

van Boxtel, C., van der Linden, J., & Kanselaar, G. (2000). Collaborative learning tasks and the elaboration of conceptual knowledge. *Learning and Instruction,* 10, 311-330.

van Boxtel, C., van der Linden, J., Roelofs, E., & Erkens, G. (2002). Collaborative concept mapping: Provoking and supporting meaningful discourse. *Theory into Practice,* 41(1), 40-46

Vattam, S., Kramer, C., Kim, H., & Kolodner, J. L. (2007). Effects of technology-based support for explanation construction on learners' discourse during design-based learning in science. In C. Chinn, G. Erkens, & S. Puntambekar (Eds.), *Mice, Minds, and Society: Proceedings of the Computer Supported Collaborative Learning Conference 2007*, Vol. 1, 740-749. Rutgers, NJ: International Society of the Learning Sciences.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds. and Trans.) Cambridge, MA: Harvard University Press.

Weinberger, A., Stegmann, K., & Fischer, F. (2006). Scripting Online Discussions: Effects on Argumentative Discourse, Cognitive Processes, and Knowledge Acquisition. In A. Weinberger, D. Clark, G. Erkens, V. Sampson, K. Stegmann, J. Janssen, J. Jaspers, G. Kanselaar, & F. Fischer (Eds.), *Argumentative Knowledge Construction in CSCL*: *Proceedings of the Seventh International Conference of the Learning Sciences* (ICLS 2006), 1094-1100. Bloomington: International Society of the Learning Sciences.

* The first two authors listed contributed equally to the writing of this paper and appear alphabetically.

Acknowledgements

CoMPASS research is supported by the National Science Foundation: IERI #0437660 & Institute of Educational Sciences: CASL #R305A08050

Weblines: Enabling the Social Transfer of Web Search Expertise using User-Generated Short-form Timelines

Neema Moraveji, Salman Ahmad, Chigusa Kita, Frank Chen, Sep Kamvar, Stanford University
Email: neema@stanford.edu, saahmad@stanford.edu, ckita@stanford.edu, fxchen@stanford.edu, sdkamvar@stanford.edu

Abstract: Web search encompasses more than fact retrieval; it is a primary entry point for learning. Exploratory search tasks are attempts at such learning and require cognitive, strategic, and interpretive work from the user. The pathways of such searches are likewise complex and nuanced. The present study attempts to enable the human work that goes into conducting exploratory searches to be efficiently captured and transmitted to other learners. By this method, web search expertise can transfer socially and implicitly between users instead of developing individually or through directed learning. The system we deployed uses an existing metaphor, the timeline, to structure insights from searches. We refer to these semantically meaningful representations as 'weblines'. We deployed a live system to 81 users in three user populations. The resulting weblines were delineated into four types. Successful weblines were those that participants used to iteratively reflect upon the insights of their searches.

Introduction and Motivation

Web search has emerged as a primary entry point to acquiring, confirming, and analyzing knowledge. The study of search has grown to encompass more than looking up discrete facts (e.g. In what year was the San Francisco fire?) and into the informal and formal learning of complex topics (e.g. Why did my local public swimming pool close and what can be done about it?).

Moving beyond the traversal of digital libraries and into general knowledge acquisition it is a skill that crosses the boundaries between formal and informal learning environments and across cultural and socioeconomic strata. While we tend to think of web search as 'information retrieval' rather than learning, much learning in information societies takes place during web search.

Information-processing theory posits that the formal mental operations required to conduct purposeful and effective searches (e.g. as required to define abstract keywords) emerge at the corresponding developmental stage. We view the logical steps in web search as a partial and abstracted representation of the user's cognitive process. By representing such processes we allow them to be transferred between participants via cognitive apprenticeship, a process by which learners can model accurate problem-solving skills by inspecting the cognitive processes of instructors (Collins 1987).

Insights: The Products of Exploratory Web Search

Researchers are actively attempting to unpack search into different types that vary in complexity. An oft-cited theory is the distinction between lookup, learning, and investigation tasks (Marchionini, 2006) (see Figure 1) while another attempts to use navigation characteristics such as 'search moves' to describe the complexity of the search task (Aula, Russell, 2008).

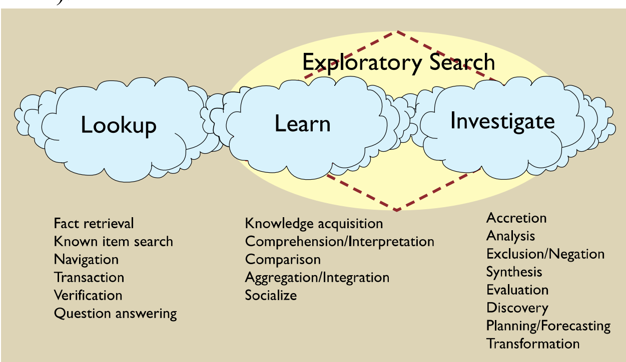


Figure 1. Web search is a broad area. The research community is growing to consider Exploratory search as being of utmost importance as it relates to higher-order tasks and processes. The results from such searches are ostensibly of greater value to the searcher and others (from (Marchionini, 2006)).

Improving system and user efficacy for exploratory search tasks has emerged as a paramount problem of human-information interaction research (Marchionini, 2006). Currently, exploratory search success depends largely on the cognitive and interpretive work of the user rather than retrieval algorithms. This human work increases the potential value of the results to the searcher and potentially to others. That work also contains cognitive processes and interpretations that are likely to be instructive to other learners about the topic at hand or the search process itself.

The results of exploratory searches are also different from those of lookup tasks because the insights gained often come from connections found across multiple webpages. This leads to a problem the present work aims to solve: how can we support users to share a series of related URLs that led to the insight in question?

For lookup tasks, the problem of sharing results has been largely solved using ‘link-sharing’ mechanisms common to social networking status updates and micro-blogging where a single URL is qualified with an optional textual description or conclusion. However, the problem is not solved for more complex and multi-step queries where multiple pages and queries are necessary to qualify and communicate the insight. As Hamming was paraphrased in (Gersh, *et al*, 2006), “the purpose of exploratory search is insight, not data.” (The original quote was “The purpose of computing is insight, not numbers.”)

We refer to the results of a complex search as ‘insights’. Insights are the result of “a story implied by relationships among discovered information items” (Gersh, *et al*, 2006). In this study we present work conducted to capture and represent the insights from complex search for later individual use as well as the social transmission of knowledge. We narrowed our focus to results from news-related searches as these encompass common research methods as well as being of social interest.

After first motivating the study of sharing insights, we present a summary of related research on the topic. This is followed by a description of our iterative, user-centered design process that draws upon principles from micro-blogging, print media, and social media. We present results using the deployed system with 81 unique users and 195 shared insights. We studied the nature of these insights and what differentiates them. Our results have implications for the study of learning technologies concerning web search expertise and knowledge-sharing systems within classrooms and organizations.

Related Work

A comprehensive and nuanced understanding of the development of search expertise (and “online reading”) is emerging in the literature (Leu, Zawilinksi, 2007). Formulating queries is complex task due to its reliance upon abstraction, vocabulary, domain knowledge, and grammar. Despite this complexity, query formulation is often the focus of explicit instruction because it is so central to Web search practice (Lazonder, 2003).

Prior work studying the habits and behaviors of students of various ages using the Internet in directed or self-directed tasks have associated novice searchers with issuing vague queries (Bilal, 2000), focusing on surface site features (Coiro, 2003), and limited confidence about and awareness of Web resources (Druin, 2010).

Numerous search curricula have been proposed though none are yet established as standards. Instructional guidelines have been proposed for teaching search skills and creating relevant learning environments as reviewed in (Lazonder, 2003). These approaches have focused on direct learning, not social, implicit learning.

The concept of imitating a skilled practitioner is perhaps the oldest and most natural form of social learning (Lave, Wenger, 1991). However, it is not clear if imitating an expert’s queries or navigational choices directly enhances a searcher’s skill set. Rather, it is the process underlying a search that is useful to imitate. This is because information needs vary from moment to moment, making the user’s conceptual model of search paramount. In postulating that these processes may be socially transferred, we draw upon cognitive apprenticeship (Collins, 1987), which posits that learners can imitate experts by deriving components of their thought processes, not just their behavior.

Sharing the Web Search Process with Others

Another branch of research in this area has been on how search processes can be represented and shared. Indeed, Vannevar Bush conceptualized “trailblazers” whose primary function in society would be to establish paths through hypertext collections for the purpose of serving future searchers (Bush, 1945). High information-density visualizations of web navigation processes have traditionally been used to study information-seeking behavior.

The work of Lin, *et al*, on visualizing search processes (Lin, 1991) (see Figure 3) led to representations like PadPrints (Hightower, *et al*, 1998) for search behavior analysis and navigation flow maps (Lin, Tsai, 2005) for enabling search instructors to understand student search behavior.

Academics have also used graphical representations of search processes for learning purposes. Twidale, *et al*, attempted to apply such representations to support collaborative search explicitly for the purpose of improving expertise (rather than improving instruction) in searching structured digital libraries with Ariadne (Twidale, 2005). These and related systems are useful for instructors and curious searchers in formal learning environments. However, they require learners be explicitly interested in improving their search expertise. We propose drawing learners into the search process of another expert through a compelling narrative.

A number of commercial applications are enable publishing and sharing of navigation histories or a collection of links about a topic. They include TrailFire.com (now defunct), PearlTrees.com, and Google Bookmarks. As far as we have seen, they have not differed significantly from any prior academic projects.

Sharing the Results from Search and Analysis

Gersh, *et al*, supports the sharing of insights in the intelligence community (Gersh, 2006). Their tools aim to share "chains of data, evidence, hypotheses, and other constructs, in which a collection of lower-level information supports a statement (hypothetical or real) at some higher level of organization or abstraction".

What remains understudied is how the results from exploratory search can be structured in a lightweight and flexible manner that can be shared to aid social learning and lead to well-acknowledged search expertise exemplars and practices. Such a representation could improve knowledge-transfer within organizations, classrooms, or in public.

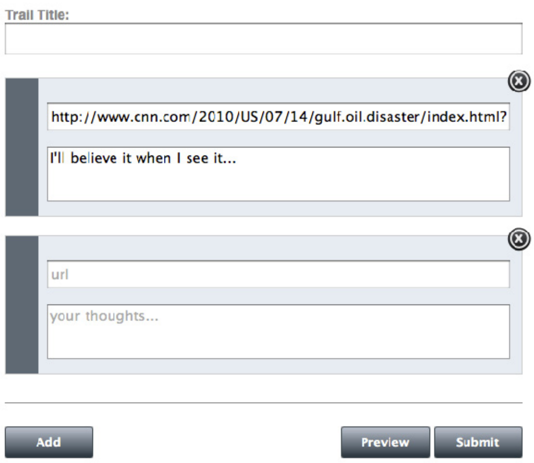
A User-centered Design Process for a Tool to Share Web-based Insights

We set out to design a system that allows users to create a single representation from a small collection of URLs that result from a complex or exploratory search. We expected the representation to consist of four components: (1) a collection of URLs, (2) the relevant portion of each URL, (3) a semantically cohesive title or meta-description, and (4) a visual representation that makes this collection easily and quickly understandable.

Though the system is designed to support social learning of web search expertise, its ostensible goal was to share insights from exploratory web search. This section describes how the final system, LineHive.com, evolved through three rounds of user testing. The final design (see Figure 4b) uses an embeddable timeline representation to aid in describing a narrative and constrains the set to ensure all items are always visible simultaneously.

Connecting a Collection of Related URLs

Based on existing practices, the initial design approach was to allow a user to declare a collection of individual URLs in a manner that closely resembles existing link-sharing practices on micro-blogging services and social networks. Such practices support sharing results from lookup tasks, *i.e.*, a single URL and an optional user-generated caption. A title was added to allow the user to describe the collection, as shown in Figure 2.



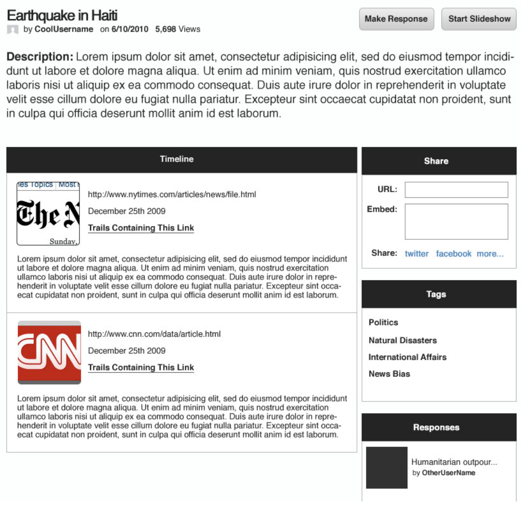


Figure 2. Our initial approach allowed a user to define a title for a collection of URLs. Each URL was then added manually with a subtitle (labeled, "your thoughts…").

Figure 3. Our second iteration was a weblog view of a URL sequence. A rich description is included to support writing of a narrative, an 'embed code' field is included.

We asked three university students to create a collection of links using a medium-fidelity prototype of this system. The primary feedback we received was that commentary or a preface was necessary to describe the sequence of URLs. This initial prototype did not feel sufficiently easy to create and share. We noted how disjointed the items were from one another. We reflected upon how, in contrast, user-generated blog posts can contain multiple URLs yet still feel cohesive because a narrative connects them.

A Weblog View of a URL Collection

Figure 3 depicts the next iteration of the interface, which added a description under the title for narrative or explanatory text. The items were laid out as individual, short weblog entries, using a well-known online metaphor (*i.e.* blogs). We also added support for reader comments and related 'trails' (as they were then called).

To support the transfer of these trails as units, we added support for embedding them as an HTML widget (as is popular with YouTube videos).

Our team conducted informal feedback sessions with several adults and found a number of problems. The description field would likely prompt users to type information, making creation prohibitively laborious. The sequence or relationship between individual URLs in the collection had to be described textually by the author, resulting in even more text. This extra information would make it time-consuming to read and result in a reduced ability to be treated as a single structural unit and theme. Similarly, the vertical scrolling required to view more than two links meant the reader would never see the sequence in its entirety.

Our approach to remedy these issues was to employ a metaphor that would constrain the URL collections but enable fast inference of the relationship between each URL. This would reduce the amount of text required to be written.

Weblines: Using Timelines to Represent Insights from Web Navigation

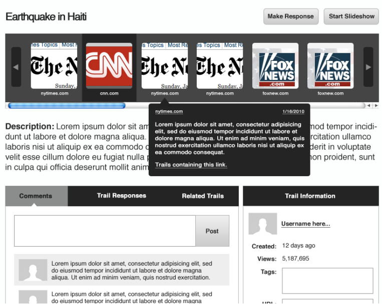
Using timelines was an attempt to turn the URL collection into a URL *sequence*. The timeline is a well-understood metaphor that includes its own narrative: setting, apex, and resolution along a temporal dimension with any units. Timelines also connote a story rather than simply a logical collection; timelines are used to outline stories and they make regular of use of thumbnails, a useful visual element for digital media.

We refer to these timelines as *weblines* because they are timelines about web traversals meant to guide another's sense-making process on the web. That is, individual items must link to URLs, they cannot contain arbitrary information or events. This constrains what weblines can be used to express but enable readers to interrogate and re-interpret the source URLs.

Following Western tradition, the timeline is presented horizontally. This was also done to ensure the sequence felt like and could be embedded in other webpages more easily. The authoring experience required the author to add a URL for the webpage and one for the thumbnail. User-defined subtitles of each URL appear on mouseover as tooltips when the user's mouse hovers atop the thumbnail.

There are two potential dates to use for each item in a webline: publish date and user access date. Our initial implementation ignores the latter and attempts to parse the former from the source URL but makes it user-configurable.

The result was a highly visual sequence of images meant to tell the story using the imagery alone. In feedback using a high-fidelity prototype with five users, we found it difficult to create a sequence whose images alone would be communicative. Figure 4a shows what would be displayed if you viewed one webline alone. The user's mouse is hovering on the third item from the left. Figure 4b shows a webline embedded in another webpage (without reader comments and metadata).



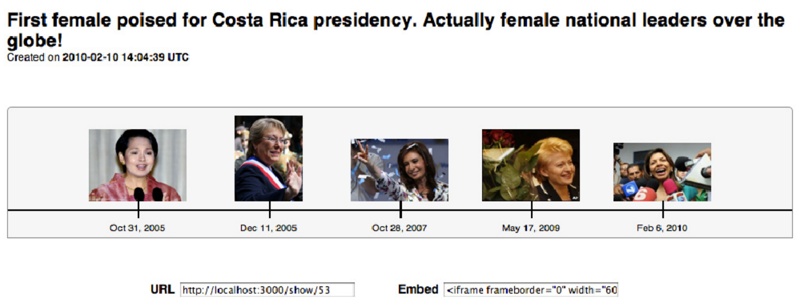
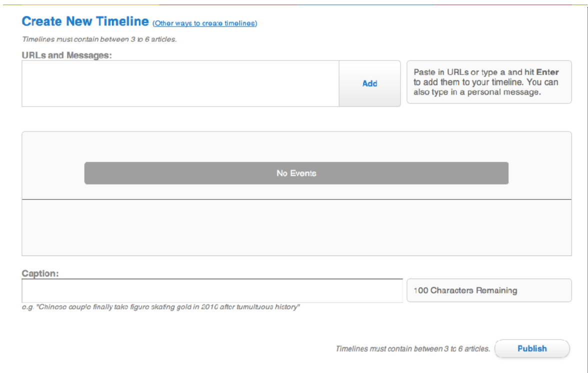


Figure 4a. A webpage dedicated to displaying a single webline. Thumbnails depict the items in the collection of URLs and a textual description describes the narrative. The reader here is hovering their mouse atop a thumbnail to reveal the subtitle.

Figure 4b. A webline embedded in another webpage hides the narrative description.

Though the initial impression of the page was less loquacious and more visual than the previous iteration, hiding the captions in tooltips meant the users co-opted the description fields to describe the narrative, again creating an abundance of text. The distinction between webline title and item description prompted users to type short titles (3-7 words) and very long descriptions. The goal for the next revision was to make the narrative immediately apparent by complementing the thumbnails with text and reducing reliance on the description field.



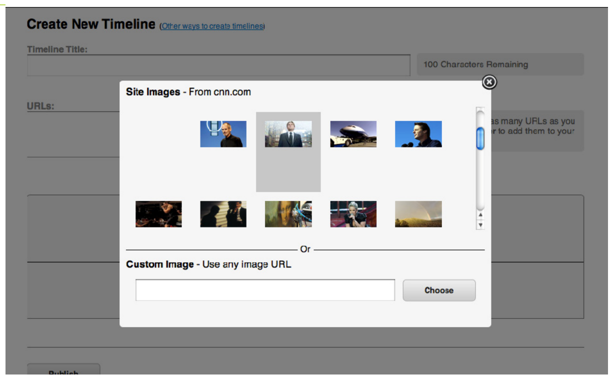


Figure 5a. The WYSIWYG authoring interface has essentially two elements: a textbox to paste URLs into and the resulting timeline below it. A single box enables the author to add items to the webline. As dates are changed of individual items, their order animates to the new sequence.

Figure 5b. After entering an item URL, the user can elect to refine the choice of a representative thumbnail. A list of thumbnails is presented from the URL in question and the user can paste any image URL as well.

WYSIWYG Authoring of Weblines

A system supporting webline authoring, reading, and embedding is available at LineHive.com. The weblines authoring interface requires the user to paste desired URLs into a textbox and press [Enter], whereupon they are added to the webline in WYSIWYG fashion. Hovering over individual items allows the user to refine the date or change the thumbnail. Crucially, as dates on individual items are changed, their position on the timeline is updated immediately, animating to the new position.

We also made timeline creation simpler by automatically downloading a thumbnail from the item's URL (an idea observed in Facebook). Because automatically selecting a representative image is error-prone, the user can refine the choice or enter the URL of a desired thumbnail image (Figure 5b).

Three types of time-scale representations were explored: *relative* (the distance between the items is proportionate to the duration between them), *fixed-interval* (all items are displayed at a fixed distance apart – see Figure 6a), and *fixed-interval clusters* (visual indicators are placed on the timeline to represent significant durations – see Figure 6b). To reduce information complexity, we sacrificed temporal accuracy and used fixed-interval clusters.

The description was eliminated and renamed to 'caption' to prompt a descriptive title. It was also moved below the timeline, resembling a caption to an illustration, to ensure the author is sufficiently descriptive.

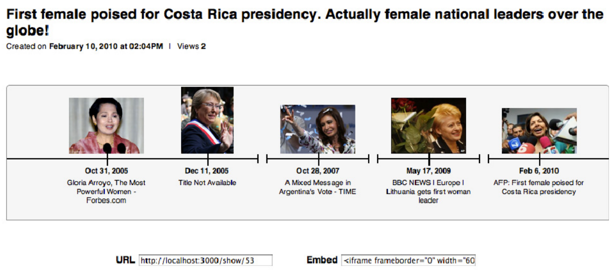


Figure 6a and b. These two weblines illustrate the difference between fixed-interval and fixed-interval with clusters types of time-scale represntations. In (a), the items are placed at fixed intervals regardless of the duration between them. In (b), visual 'time gap' indicators indicate relatively large durations.

Sharing Weblines by Embedding or Hyperlinking

The sequence of webpages that a webline represents is meant to be treated as a unit of information complementing a greater point or standing alone. Each webline has a short URL that can be micro-blogged and can be embedded as a widget to complement articles or blog posts.

Pilot Study: 60 Collocated Participants in a Classroom Setting

We took a high-fidelity prototype of the design to two 9th grade public high school journalism classrooms in an affluent area where students were writing reviews of books by conducting Internet research. They were asked to create timelines to complement their reviews.

The function and form of the caption was effective at eliciting descriptive but not loquacious titles. Though we expected these timelines to be created quickly, the students voiced their opinion to login, save drafts, and revisit the timelines they created. In effect, the timelines became valuable narrative representation of student thinking and research. They both took longer to create and were more valuable to authors and peers than expected.

Frequently, students would create long timelines with over 10 items that required significant horizontal scrolling, an undesired consequence. They would simply paste links from any URLs they visited during their research, without regard. These quickly became unwieldy and lost their narrative thread.

Constraining Volume and Length to Form Short-form Timelines

Our team inspected a number of print periodicals for examples of timeline narratives. We observed that short-form timelines were used frequently to complement articles or even to stand on their own. These short-form timelines were constrained in both volume (number of items) and length (characters of text). Readers are accustomed to them and they drew upon the same principle that has led to the success of micro-blogging: constraint. We restricted the maximum number of items to 6 and amount of characters to 200 for each subtitle, and the title to 100 characters (to make room for a URL that could then be micro-blogged).

By restricting the volume and length of our timelines would both eliminate horizontal scrolling and, ideally, ensure tighter narratives. By reducing the number of options, authors would have to be selective in their choice of URLs. All items would be visible on one screen at one time, requiring no scrolling whatsoever. This became a guiding principle in the system's design.

Evaluation

Our research question then became, "What characterizes the nature of different weblines and how does the user's exploratory search process inform this nature?"

Participants, Dataset, and Methodology

We deployed LineHive.com publicly and worked closely with three groups of participants to collect data on how it was used and the nature of the resulting timelines over the course of three months. The first group N1, was a classroom of 30 9th grade journalism students. The second group, N2, consisted of 29 journalists for an online magazine who work remotely from one another and conduct weekly teleconferences (and occasional in-person meetings) that our team became part of. The third, N3, was a group of 22 students in a university course on social technologies. The students had weekly reading assignments and, each week, were given the option of writing an essay or creating a webline. In either case, they were shared publicly with the class and on Twitter. In total, there were 81 unique users of the tool. In total, 353 weblines were analyzed (including the research teams' own examples). We collected and analyzed the weblines created by all these participants.

We used a grounded theory (Glaser, 1992) approach to evaluate how users expressed their search insights. We pruned a portion of weblines that were not suitable for analysis. The pruning process was as follows. First, we first omitted duplicates and some improper contents such as racist account (which were likely created by arbitrary web users). Second, we omitted 'test' weblines. Through this pruning process, 252 timelines remained to be analyzed.

We then categorized weblines based on communicative intent. With these initial categories, we re-examined the 252 weblines and refined the categories again. Two category types emerged: "sequence" and "argument." We then pruned items that our team had contributed, resulting in 195 weblines. Weblines in the sequence category pull together links in order to *describe* an event (e.g. "Major events of the Obama campaign") or steps in a process while argument weblines are where the author is using links to make a logical argument or statement (e.g. "How social media changed youth radio").

Results

Among the 195 weblines created, 26 lines deal strictly with chronological report of a current issue (13.3%) (e.g. Figure 7a). In creating these narratives, authors had to choose the most relevant facts to form a sequence of events on an issue. However, chronology of an issue sometimes has longer time span as shown in Figure 7b. Figure 7a was clearly stimulated by the Olympic Games but is focused on its historical aspect. There are 61 historical weblines of this kind (31.3%).

A portion of the weblines go beyond just convey a chronology as Figure 7c illustrates. In this case, the caption delivers creator's argument, and the links are selected as to support the argument, at the same time the evidences became more comprehensive when it is in chronological order. 22 timelines are fallen into this category (11.2%).

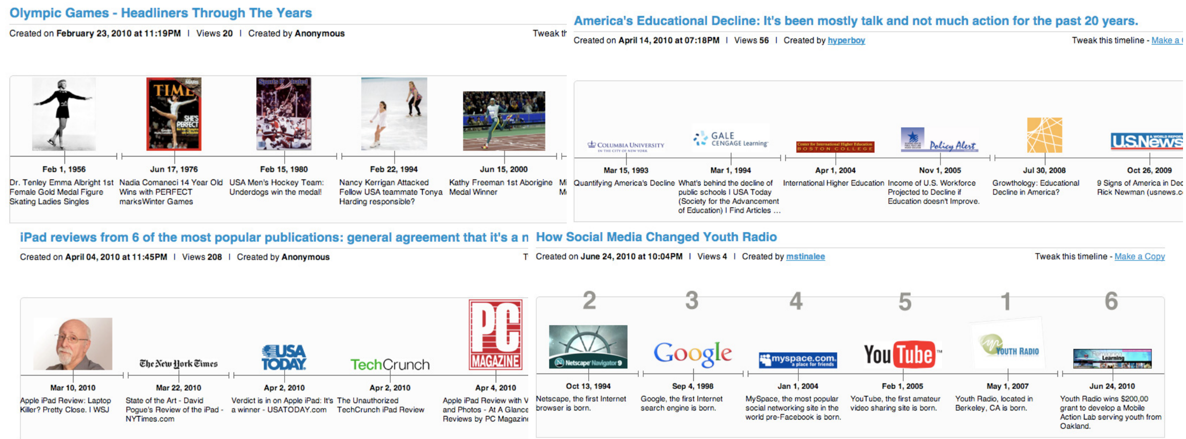


Figure 7a-d, clockwise from top-left. (a) This webline illustrates a chronological timeline that does not present an argument, *per se*. (b) This webline combines a chronology with an argument. (c) The result of a contextual interview we conducted, this webline both makes an argument and presents a chronology of a topic. (d) This webline does not present a chronology of an event but rather a collection of weblines. The numbers on this latter webline indicate the order in which they were added by the author.

In a majority of the weblines that illustrate a well-constructed argument, the chronological order of the URLs is not paramount, as shown in Figure 7d. We labeled this category, which contains 43 weblines (22.0%), "argument by example."

We then created a quadrant graph of the weblines with the two dimensions, plotted in Figure 8 (with counts for each quadrant). One third of timelines made a clear argument but it is not clear whether this type of argument is formed during the search process or beforehand.

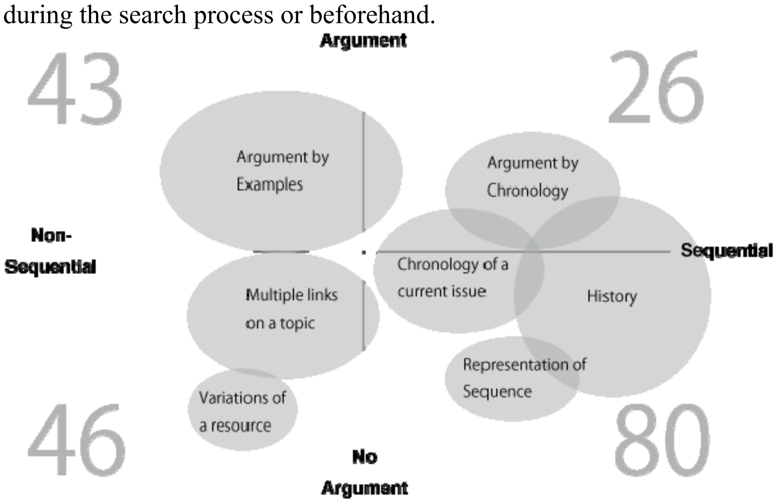


Figure 8. Weblines made using the tool fell into four quadrants along two dimensions: Sequence and Argument. The numbers represent counts. Category descriptors on the circles represent sub-categories.

Contextual Interview with an Expert Webline Creator

We conducted a contextual interview with a participant who created multiple high-quality weblines to see how such insights are developed and transferred into weblines. We tasked her with making a webline related to her expertise in youth's use of social media. We asked her to think-aloud (Lewis, 1982) through this process.

The participant spent 48 minutes talking. We attribute this relatively long duration to the think-aloud and interjecting we initiated. During this time, she changed URLs 58 times, 11 of which were searches. The outcome is as Fig. 7d. The number on each link indicates the order in which that link was added to the webline. This timeline can be plotted in the overlap area between "History" and "Argument by Chronology" in Fig. 8. The webline caption was finalized at the end. Her initial intention evolved significantly over the course of creation. She frequently mentioned the remaining number of link spots available and ran up to this constraint repeatedly. Several times throughout the process, she looked at what she had so far and the forced temporal sequence made her re-evaluate her goal and the most interesting insight to share with the webline.

Arguments shown in weblines may not be directly attributed as an initial insight but the tool galvanizes the exploratory search and insights from the search can be expressed as a selection of the items. In effect, a number of non-obvious insights can support weblines that are ostensibly made to be arguments or chronologies. Weblines effectively support not only the transfer of insights but also an insightful search process itself.

Discussion and Next Steps

We learned how the exploratory search process can be represented using a metaphor common to print design: the short-form timeline. By translating search skill development into a constructionist (Harel, Papert, 1991) storytelling tool, authors can create weblines individually or collaboratively. Future studies will determine if authors develop exploratory search skills by ensuring their search process develops a narrative arc and, for readers, because exploratory search process is represented by an engaging narrative. It is core to the design that each webline is automatically published in a user's 'stream' so that public sharing is the default.

One reason users iterate on their weblines is because they re-organize themselves based on each link's publication date. This forces users to re-evaluate their choices and address gaps. By having at least one organizing principle (*i.e.* time), the system enforces a level of coherency in the user's findings.

Our representation does not illustrate the author's entire process, as prior work did, but rather the resulting insight. Thus, the trade-off is that authors are motivated to search enough to produce an insight but that this process is more opaque to readers. Future work might reveal this process upon request.

This study has implications for how exploratory search systems are conceived. Currently, search interfaces are designed for solitary use and their results are of use to the user only; sharing is a secondary consideration. It is interesting to investigate search engines as inherently social systems such that the act of using it was to create knowledge for oneself and others. By creating knowledge one knows others will consume, constructionist theory posits that one creates higher-quality artifacts and conducts more insightful searches.

References

Aula, A., Russell, D. (2008). Complex and Exploratory Web Search. *NSF: Information Seeking Support Systems*. UNC Chapel Hill.

Bush, Vannevar. (1945). "As We May Think." *The Atlantic Monthly*.

Bilal, D. (2000). Children's use of the Yahooligans! Web Search Engine: I. Cognitive, Physical, and Affective Behaviors on Fact-Based Search-Tasks. *J American Society for Information Sciences*, 51.7, 646-5.

Coiro, J. (2003). Rethinking comprehension strategies to better prepare students for critically evaluating content on the Internet. New Eng Reading Assoc J, 39, 29-34.

Collins, A., Brown, J.S., Newman, S.E. (1989) Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics. Knowing Learning and Instruction, 8.1, 453-94.

Druin, A., Foss, E., Hutchinson, H., Golub, E., Hatley, L. (2010). Children's roles using keyword search interfaces at home. *Proceedings of CHI 2010*.

Gersh, J., Lewis, B., Montemayor, J., Piatko, C., Turner, R. (2006). Supporting insight-based information exploration in intelligence analysis. Commun. ACM 49, 4, 63-68.

Glaser, B. (1992). Basics of grounded theory analysis. Mill Valley, CA: Sociology Press.

Gupta, N.K., Rosé, C.P. (2010) Understanding Instructional Support Needs of Emerging Internet Users for Web-based Information Seeking. *Journal of Educational Data Mining*, 2 (1), 38-82.

Harel, I., Papert, S. (1991). Constructionism. Ablex Publishing.

Hightower, R. R., Ring, L. T., Helfman, J. I., Bederson, B. B., Hollan, J. D. (1998) Graphical multiscale Web histories: a study of padprints. *Proceedings of Hypertext*.

Kuiper, E., Volman, M., Terwel, J. (2005). The Web as an Information Resource in K-12 Education: Strategies for Supporting Students. *Rev of Educ Res*, 75.3, 285-328.

Lave, J. & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge Univ Press.

Lazonder, A. (2003). Principles for designing web searching instruction. *Educ. & Info Tech.*, 8.2, 179-93.

Leu, D.J., Zawilinski, L. (2007). The New Literacies of Online Reading Comprehension. *JNERA*, 43.1, 1-7.

Lewis, C. H. (1982). Using the "Thinking Aloud" Method In Cognitive Interface Design. IBM TR RC-9265.

Lin, X., Liebscher, P. Marchionni, G. (1991). Graphical representations of electronic search patterns, *J Amer Soc for Info Sci*, 42.7, 469-78.

Lin, C., Tsai, C. (2005). Navigation Flow Map Method of Representing Students' Searching Strategies on the Web. *Proceedings of World Conf on Educational Multimedia, Hypermedia and Telecommunications*.

Marchionini, G. (2006). Exploratory search: from finding to understanding. Commun. ACM 49, 4, 41-46.

Twidale, M. B. (2005). Over the Shoulder Learning: Supporting Brief Informal Learning. *JCSCW*, 14.6.

Acknowledgements

Chigusa Kita's research was supported by Kansai University's Overseas Research Program for the year of 2010. We thank David Zhang and Amanda Li for interface design support and the reviewers for valuable feedback. We thank Roy Pea for valuable guidance.

Facilitating Web Design Skills through Online Design-Based Learning: The Case of Collaboration Scripts and Incomplete Concept Maps

Walid El-Refai, Ingo Kollar, Frank Fischer, Ludwig-Maximilians-University (LMU) Munich, Institute of Educational Psychology, Leopoldstr. 13, D - 80802 Munich
Email: walid.el-refai@psy.lmu.de, ingo.kollar@psy.lmu.de, frank.fischer@psy.lmu.de

Abstract: Web design skills are an important component of media literacy. The aim of our study was to promote university students' web design skills through online design-based learning (DBL). Combined in a 2x2-factorial design, two types of scaffolding were implemented in an online DBL environment to support the students through their effort to design, build, modify, and publish web sites on processes and outcomes measures, namely collaboration scripts and incomplete concept maps. The results showed that both treatments had positive effects on domain-specific knowledge acquisition and on content-related discourse quality and collaboration skills in a subsequent discussion. There was synergism between the two scaffolds in that the combination of the collaboration script and incomplete concept maps produced the most positive results. To be effective, online DBL thus needs to be enhanced by appropriate scaffolds, and both collaboration scripts and incomplete concept maps are effective examples.

Web Design Skills as a Component of Media Literacy

Media literacy has become an important competence in our society (Piette & Giroux, 2001). It concerns the competence to both critically analyze media and to create messages in a wide variety of forms (e.g. print, audio, video, and multimedia; Hobbs, 1998). Designing web content is one important component of media literacy, which however receives little attention in college and university education as reflected by empirical evidence showing that web design skills of university students are not very high (e.g. Shannon, 2008; Wallace & Clariana, 2005). In deed, web design is a challenging and complex process that requires learners to build complex web sites involving hyperlinked documents, different navigation systems, and social information spaces as well as to account for the possible needs, purposes, and abilities of different users (Spyridakis, Wei, Barrick, Cuddihy & Maust, 2005). In this study, we investigate how the acquisition of university students' web design skills can be supported in CSCL environments.

Facilitating Web Design Skills through Online Design-Based Learning

One promising approach to support students' acquisition of web design skills is Design-based Learning (DBL). DBL has been described as a combination of problem-based learning and inquiry learning and has gained wide-spread attention in the Learning Sciences (e.g. Fortus, Dershimer, Krajcik, Marx & Mamlok-Naaman, 2004; Kolodner, 2002). In DBL, students engage in design tasks such as building and creating computer programs, models, projects, etc. Engaging in such design processes is supposed to support learning of content knowledge as well as the acquisition of social and communicative, inquiry-related skills (e.g., communication, collaboration and problem-solving skills; see Kolodner, 2002). The design task is defined as an ill-structured problem that encourages students to engage in design-based discussions, which involves asking questions, searching and giving answers, evaluating answers and engaging in discussions about the content. Educational research has developed quite a number of different DBL-approaches such as Learning by Design (LBD; Kolodner, 2002), Design-Based Science (DBS; Fortus et al., 2004), and Engineering Competitions (Sadler, Coyle & Schwartz, 2000). In LBD, students go through two cycles of activities. One cycle concerns the *design* of the artifact (e.g., designing and building a miniature car that goes from one end of the classroom to the other), while the other focuses on the *investigation* of the designed artifacts by means of controlled experiments. In both cycles, activities on different social levels (individual, small groups, and plenary) are realized to help students acquire science knowledge and skills and to benefit from each others' comments and suggestions. We investigated if DBL can successfully be transferred to online learning and be used to facilitate the acquisition of web design skills. However, given the complexity of DBL, we were interested in whether students can benefit from further instructional support measures such as incomplete concept maps and collaboration scripts.

Supporting Online Design-Based Learning through Collaboration Scripts and Incomplete Concept Maps

As described, DBL requires students to engage in complex design processes and challenging investigation processes. In addition, when students work in small groups, they need to engage in high-level collaboration processes that seldom happen spontaneously (Cohen, 1994). Therefore, additional support during online DBL seems to be promising. There are at least two categories of scaffolding that have a certain potential to improve online DBL: (a) social scaffolding, that refers to the guidance and structuring of social interactions (Kollar, Fischer & Hesse, 2006), and (b) content scaffolding, that refers to conceptual support concerning the content of the task (Cox & Brna, 1995). Our study focuses on (a) computer-supported collaboration scripts as a realization of social scaffolding and (b) incomplete concept maps as a realization of content-related scaffolding.

Computer-supported Collaboration Scripts

DBL typically requires students to engage in collaborative learning activities through their investigation. In LBD, students collaborate in small groups to reach an understanding of the design task at hand, control and conduct the design process and related empirical investigations, and effectively communicate with other learning partners (Kolodner, 2002). Many studies have demonstrated that students often do not collaborate well and experience difficulties when supposed to engage in high-level collaboration processes (e.g., Cohen, 1994). If realized online, these problems may even be amplified. As a solution, collaboration scripts are considered powerful instructional interventions. They facilitate high-level collaboration processes by specifying, sequencing and distributing learning activities and roles among learners in a small group (Kollar et al., 2006). For example, Kollar, Fischer and Slotta (2007) demonstrated that collaboration scripts are able to improve collaboration processes and individual learning outcomes through structuring the interactive processes between learning partners. Moreover, collaboration scripts may facilitate communicative-coordinative processes between students and guide them through complex learning processes (Haake & Schümmer, 2003; Hoppe, Gaßner, Mühlenbrock & Tewissen, 2000). With respect to knowledge acquisition, there is evidence that collaboration scripts can positively affect especially the acquisition of more domain-general knowledge, but if properly designed also may facilitate the acquisition of domain-specific knowledge (Weinberger, Stegmann & Fischer, 2010). Based on this, we expected that implementing a collaboration script in an online DBL environment would lead to more domain-specific knowledge concerning learning content and domain-general knowledge on collaboration than unstructured collaboration.

Incomplete Concept Maps

In online DBL, students have to acquire science concepts and realize their mutual relations, after that connect them with ideas about research and design, and finally organize their ideas for addressing the challenge (Vattam & Kolodner, 2006). To help achieve this, shared external representations that provide students with visual representations of the to-be-learned content have been demonstrated useful (e.g., Suthers, 2003). In online DBL, one type of shared external representations that may be useful are concept maps, which are visual representations that graphically depict relations between concepts (Baker, 2003; Cox & Brna, 1995). Studies on the use of concept maps and graphical organizers indicate positive effects on comprehension of content (e.g. Romance & Vitale, 2002), intensive discourse about content (Fischer, Bruhn, Gräsel, & Mandl, 2002), and content quality of discussions (Jacobson & Levin, 1995; Suthers, 2003). One strength of using concept maps lies in the graphic representation to visualize and manage domain-specific knowledge necessary for the design task (Tergan & Keller, 2005), which enables students to understand and remember complex information and abstract concept relationships (Armstrong, 2003) as well as to organize and communicate their ideas and thoughts (Cox & Brna, 1995; Jonassen, Beissner, & Yacci, 1993). Even more, concept maps that involve missing concepts and relationships can increase students' activities for seeking missing information, which in turn may positively affect learning outcomes (Baker, 2003). Thus, in this study we used incomplete concept maps and implemented them in an online DBL environment. We expected that incomplete concept maps would increase students' discussions about content and thus lead to higher levels of domain specific knowledge and collaboration skills than unsupported DBL.

Research Questions and Hypotheses

In this study we examined to what extent a collaboration script and incomplete concept maps as well as their combination affect (1) the *acquisition of domain-specific knowledge* and *domain-specific skills* related to the design and building of web sites and (2) the content-related *discourse quality* and *collaboration skills* shown in a subsequent collaborative transfer task. We expected positive effects of both treatments on both kinds of outcomes and hypothesized the combination of collaboration script and incomplete concept maps to lead to the highest results in the post tests.

Method

Participants

100 students from the Educational Technology Department of Tanta University (Egypt) participated in the study in winter term of 2009. 15 participants were male, and 85 participants were female (M_{Age} = 20.00, SD = .71). Their task was to design, build and publish tourist web sites about Egypt by aid of the software “FrontPage”.

Design

An experimental 2x2-factorial design was established with the collaboration script (without vs. with) and incomplete concept maps (without vs. with) as independent variables (see table 1). The students were randomly assigned to dyads which were then randomly assigned to one of the four conditions.

Table 1: Design of the empirical study.

		Collaboration Script	
		Without	With
Incomplete Concept Map	Without	N= 20 (10 dyads)	N= 24 (12 dyads)
	With	N= 24 (12 dyads)	N= 32 (16 dyads)

Setting and Learning Environment

The students participated in an Arabic online DBL environment, the design of which was inspired by the LBD approach (Kolodner, 2002). In each condition, the online DBL environment was divided into three sections (see figure 1). The left part of the screen (1) included tutorial videos about how to use the learning environment. The middle section (2) involved the course instructions, a news forum, and links to the learning phases, which in accordance with the course schedule appeared in a sequential order. The right section (3) included a range of communication tools (chat rooms, Wiki pages, and emails) to be used by the students and the teacher.

Figure 1. Screenshot of the Online DBL.

Each dyad had a private chat room, in which either a collaboration script, an incomplete concept map, both or none of the two were integrated (see below). Moreover, each dyad had a private Wiki page to receive comments, questions, and answers from other groups during designated sessions in which each dyad was supposed to evaluate the web sites created by one other group. The role of the teacher was to provide the students with all details about the task and instructions for each learning phase as well as technical support for the students during the learning phases.

The learning scenario included five phases. (1) The first phase started with the teacher providing the design task to the students, namely to create tourist websites on Egypt. Then, the students started their investigation by looking at both good and poor tourist web site examples, based on specific constructive standards for designing web sites, which were extracted from the literature and presented in the online environment (e.g., Harbeck & Sherman, 1999; Powell, 2001). (2) The second phase was devoted to learn about "FrontPage" software and the constructive standards for designing web sites by providing nine online tutorial lessons. At the end of each lesson, there was a small group discussion during which the different treatments were implemented. (3) Through the third phase, each dyad developed and discussed a plan for their web site and sent it to other groups before the group session. (4) During the fourth phase, each dyad built and published their web sites individually, then discussed, compared, and decided which of the two to represent their group. (5) In the fifth phase, each dyad had the opportunity to modify, rebuild, and republish their web site. In each of these phases, the students worked individually, in small groups (dyads), and in plenary sessions. The treatment was provided during the small groups discussions.

Procedure

First, pre-tests served to determine prior domain-specific knowledge and domain-specific skills related to the design and building of web sites. Subsequently, the student dyads were introduced to the online DBL environment. Afterwards, they participated in the learning phases for an overall of 41 days. In each learning phase, the students started their work individually, then conducted small group (dyadic) discussions (depending on the experimental condition with or without collaboration scripts and incomplete concept maps), and at the end of each phase, plenary sessions were conducted. In the final phase the students took two post-tests: post-test 1 was realized as an unstructured chat session to measure content-related discourse quality and collaboration skills employed in a subsequent online discussion, and post-test 2 was a test on domain-specific knowledge and domain-specific skills related to the design and building of web sites.

Experimental Conditions

The experimental design of this study consisted of the following four conditions:

(1) In the *control condition,* groups did not receive scaffolds during their small group discussions and thus led unstructured discussions using regular chat facilities.

(2) Groups in the condition *with collaboration script only* received a collection of interaction-related prompts to structure their collaborative discussions. The collaboration script involved different activities and roles for each student during the small group' discussions (see figure 2): (a) first, student 1 was instructed to ask a design-related question, (b) which student 2 then was supposed to answer. Then (c) student 1 was asked to either accept the answer with or without comment(s) or to refuse the answer with or without justification(s). After that (d) student 2 could either accept his/her partner's answer with or without comment(s) or refuse his/her partner's answer with or without justification(s). Finally (e) the discussion between dyads about the questions repeated until both students agreed on the same answer. After that the students' roles were switched to start a new cycle of the collaboration script (student 2 asks a new question and student 1 gives answer, etc.).

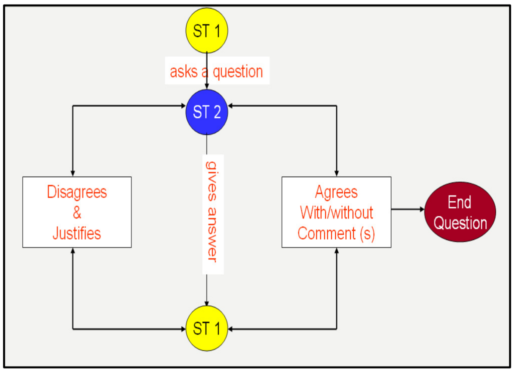


Figure 2. Visual Representation Shows Sequence of the Computer-supported Collaboration Script.

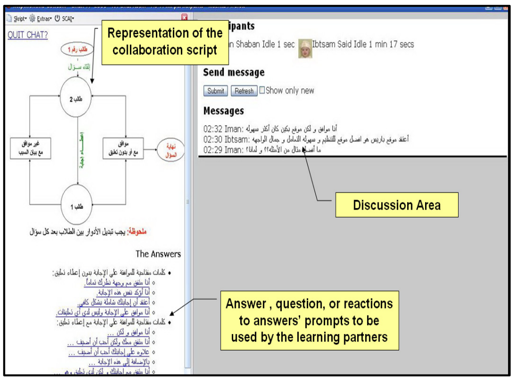


Figure 3. Chat Room Supported with the Computer-supported Collaboration Script.

In this condition, each chat room was equipped with the script (see figure 3). The right side of the chat room (discussion area) was where students could send and receive messages from their partner, while the collaboration script was located in the left side of the screen. The collaboration script section was divided into two parts: the *upper part* presented a visual representation of the script, while the *bottom part* involved specific prompts (based on the work by King, 1989) to assure a high level of discussion. There were prompts concerning questions (e.g. "Explain how…?" "What is the best…?" "Why…?"), answers (e.g. "I think the answer is…", "From my perspective…"), and reactions to answers (e.g. "I support this answer", "I agree but…", "I disagree because…"). The prompts were changed automatically according to the student's role after sending his/her message to his/her partner. In addition, the student was not allowed to write his/her message in the discussion area before selecting a prompt that was associated with his/her role.

(3) In the condition with *incomplete concept maps* only, groups were provided with 12 incomplete concept maps, one for each chat session. Each concept map involved the key concepts, principles, and propositions related to the contents of the particular learning phase, which were listed and ranked in a hierarchical order. Each level of the concept maps had the same color, shape (e.g. circles, oval, or rectangle), and size. Lines and arrows were used to indicate relationships between the concepts. We also put verbs on each arrow to clarify the kind of relationship between the two concepts. Each concept map therefore represented a

group of related concepts. In each group, some boxes and arrows were not named to evoke students' discussions about this missing information (see figure 4) and to force them to discuss all sections of the map. Students explicitly had the task to fill in blank spots in the concept map. The missing concepts were varied between specific concepts that related more to the topic (e.g. bookmarks, hotspot area, DHTML effects) as well as intermediate (e.g. hyperlinks, Marque, jump menus), and general concepts (e.g. toolbars, multimedia, and interface) that focus on specific content. The missing verbs were limited to relationships between specific concepts (e.g. verb "create" to express the relationship from "image maps" concept to "hotspot area" concept) or between specific concepts with intermediate and/or general concepts (e.g. verb "write" to express the relationship from "Marque" as intermediate concept to "Marque text" as specific concept).

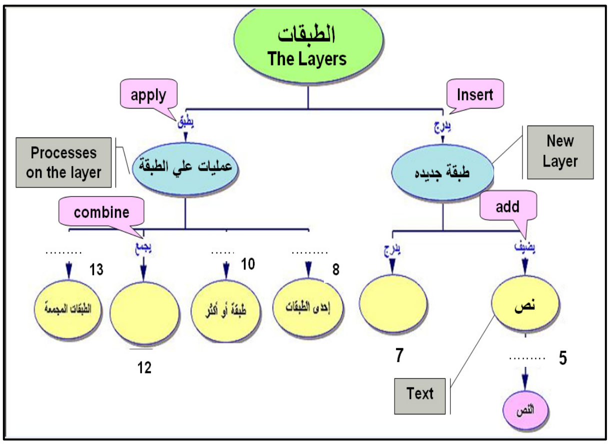


Figure 4. Screenshot of Incomplete Concept Map (Representations in Bubbles and Square Boxes Serve as Caption and were Not Presented in the Concept Map).

(4) In the *combined condition*, the students were supported with both the computer-supported collaboration script and incomplete concept maps during their small group discussions.

Dependent Variables and Instruments

Two tests were used to assess different learning outcomes: Post test 1 measured content-related discourse quality and collaboration skills in a subsequent collaborative transfer task; post test 2 assessed individuals' factual knowledge and skills on web design.

With respect to content-related discourse quality and collaboration skills, we analyzed the final discussion in each small group, which was realized without treatment. Trained coders first segmented each discourse into meaningful sentences and rated the segments (total agreement = 86 %). To assess *content-related discourse quality*, a coding scheme was developed to capture what web-design concepts were used during discussions (Cohen's κ = .71). The number of used concepts during the final chat discussion was used as indicator for content-related discourse quality. With respect to *collaboration skills*, a coding scheme was developed to identify the different kinds of questions (Cohen's κ = .97), answers (Cohen's κ = .76), and reactions to answers (Cohen's κ = .83) that were suggested by the collaboration script. After that, a principal component analysis (PCA) with all 16 variables was performed resulting in two factors which were used to describe two different dimensions of collaboration skills. The first factor included the categories "accepting answers with comment(s)" and "refusing answers with justification(s)", which we labeled as "elaborative dimension of collaboration skills". The second factor included only one variable, which was "evaluation questions". Thus, this factor was labeled "metacognitive dimension of collaboration skills".

Post test 2 measured both domain-specific factual knowledge and domain-specific skills of web design. The *factual knowledge test* consisted of 66 multiple choice questions directed towards the functionalities of FrontPage. The *application-oriented knowledge test* assessed students' Web-design skills when using FrontPage software. To that end, they were asked to perform the different functions of FrontPage and could reach between 0 (the action could not be performed) and 2 points (the action was performed correctly) for each function. Reliabilities were sufficient (α = .72 for factual knowledge test, α = .96 for application-oriented knowledge test).

Results

Collaboration Skills and Domain-Specific Discourse Quality

Table 2 presents the mean scores and standard deviations concerning collaboration skills and domain-specific discourse quality in the subsequent collaborative transfer task for the four experimental conditions.

An ANCOVA with the collaboration script and the incomplete concept maps as fixed factors and the *elaborative dimension of collaboration skills* exhibited in the final unstructured chat session as dependent variable as well as the total number of discourse segments as a covariate revealed significant and positive main effects for both treatments ($F(1.46) = 17.30$, $p < .001$, $Eta^2 = .28$ for collaboration script, and $F(1,46) = 7.43$, $p < .001$, $Eta^2 = .14$ for incomplete concept maps), and their interaction ($F(1,46) = 20.69$, $p < .001$, $Eta^2 = .32$). An analogous ANCOVA with the *metacognitive dimension of collaboration skills* exhibited in the final unstructured chat session as dependent variable revealed significant and positive main effects only for the collaboration script ($F(1.46) = 4.60$, $p < .001$, $Eta^2 = .09$). The interaction effect did not reach statistical significance ($F(1,46) = .14$, *n.s.*). An ANCOVA with the collaboration script and the incomplete concept maps as fixed factors and the *domain-specific discourse quality* as dependent variable as well as the total number of discourse segments as a covariate revealed significant and positive main effects for both treatments ($F(1,46) = 13.30$, $p < .001$, $Eta^2 = .23$ for collaboration script, and $F(1,46) = 45.11$, $p < .001$, $Eta^2 = .50$ for incomplete concept maps). The interaction effect did not reach statistical significance ($F(1,46) = .24$, *n.s.*). Due to significant effects on the Levene's test on homogeneity of variances, the effects were re-tested with non-parametrical tests (Kruskal-Wallis test and Mann-Whitney U-test), confirming the results of the ANCOVAs.

Table 2: Mean scores (standard deviations in brackets) concerning collaboration skills (elaborative and metacognitive dimension) and domain-specific discourse quality in the subsequent collaborative transfer task for the four experimental conditions.

	control group	collaboration script only	incomplete concept maps only	collaboration script and incomplete concept maps
	M *(SD)*	M *(SD)*	M *(SD)*	M *(SD)*
Collaboration skills (elaborative dimension)	- .75 *(.22)*	-.56 *(.42)*	-.47 *(.21)*	1.24 *(.79)*
Collaboration skills (metacognitive dimension)	- .75 *(.10)*	- .03 *(.88)*	- .45 *(.29)*	.83 *(1.16)*
Domain-specific discourse quality	6.40 *(2.22)*	16.75 *(9.44)*	9.08 *(.79)*	20.38 *(6.77)*

Acquisition of Domain-Specific Knowledge and Skills

For *domain-specific factual knowledge* about FrontPage software and standards for designing web sites (mean scores see Fig. 5), the results showed significant and positive effects for both the collaboration script and incomplete concept maps. An ANCOVA with the collaboration script and the incomplete concept maps as fixed factors, group membership as further independent factor nested within the experimental conditions (to account for interdependencies in the data of the two learners in a group), the post test scores in the factual knowledge test as dependent variable and the pre test scores as a covariate showed stronger effects for incomplete concept maps ($F(1,95) = 28.13$, $p < .001$, $Eta^2 = .37$) than for the collaboration script ($F(1,95) = 11.94$, $p < .001$, $Eta^2 = .20$). The interaction effect did not reach statistical significance ($F(1,95) = 2.30$; *n.s.*). Thus, the combination of collaboration script and incomplete concept maps led to the highest levels of knowledge acquisition.

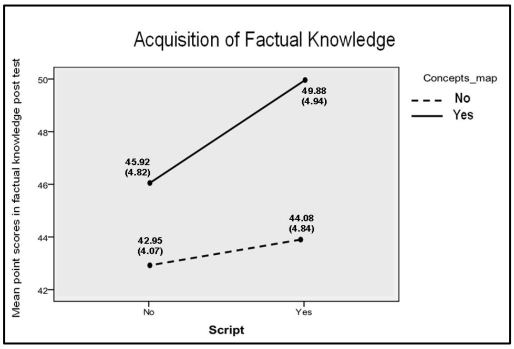


Figure 5. Mean Scores (Standard Deviations in Brackets) in the Test on Domain-Specific Factual Knowledge across the Four Experimental.

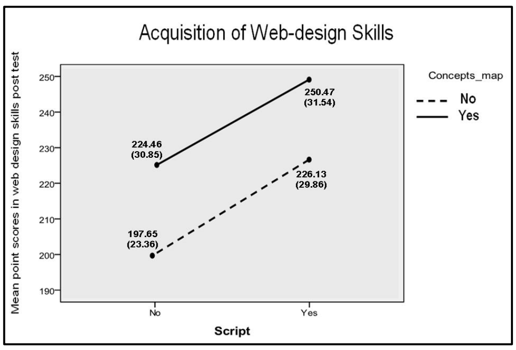


Figure 6. Mean Scores (Standard Deviations in Brackets) in the Test on Domain-Specific Web Design Skills across the Four Experimental.

A similar pattern was found with respect to the *acquisition of web design skills* (mean scores see Fig. 6). An ANCOVA with the collaboration script and the incomplete concept map as fixed factors, group membership as further independent factor nested within the experimental conditions, the post test scores in the Web-design skills test as dependent variable and the pre test scores as a covariate showed significant and positive main effects for both treatments ($F(1,95) = 21.90$, $p < .001$, $Eta^2 = .31$ for incomplete concept maps, and $F(1,95) = 27.61$, $p < .001$, $Eta^2 = .36$ for the collaboration script), but no significant interaction effect ($F(1,95) < 1$; *n.s.*). Thus again, additive effects for both treatments with respect to web design skills were found.

Discussion

Overall, our results support quite clearly our expectation that online DBL can be improved by adequate scaffolding. When considering the results in the unstructured control condition and comparing them to the other three conditions, it is confirmed that additional scaffolding can improve learners' knowledge and skill acquisition in online DBL. As our study indicates, implementing social and content scaffolds in an online DBL environment are powerful means to improve collaboration skills and content-related discourse quality in a subsequent collaborative transfer task as well as the acquisition of factual knowledge and skills on web design. Obviously, and in line with prior research (Kollar et al, 2007; Weinberger, Ertl, Fischer & Mandl, 2005), the collaboration script we used seemingly helped learners lead more sophisticated discussions, which in turn may have led to the acquisition of domain-specific factual knowledge and web-design skills. This is a very encouraging finding since many laboratory studies on collaboration scripts in CSCL contexts failed to find positive effects on more domain-specific outcomes. Perhaps, compared to early studies, the more extensive learning phase employed in this study employed may have promoted such positive effects. Yet, to test the underlying process assumptions, in-depth process analyses are necessary, which are currently under way.

In line with results of studies on the collaborative use of concept mapping techniques (e.g. Fischer et al., 2002), the incomplete concept map provided the students with effective conceptual support concerning the domain-specific content of the task. This may have facilitated the content-related, but in parts also the social-interactive level of the discussions between students (Jacobson & Levin, 1995) during their work in the online DBL environment (Tergan & Keller, 2005). Also, these higher levels of discussion and collaboration obviously may have been responsible for enabling students to acquire both factual knowledge on Web-design concepts and web-design skills. Yet again, process analyses are needed to confirm these process assumptions.

With respect to the question how collaboration scripts and incomplete concept maps work together in online DBL, we found a results pattern that can be described as a nice materialization of what Tabak (2004) called "synergistic scaffolding". On almost all dependent variables we found the combination of collaboration script and incomplete concept maps to evoke the most favorable results, going beyond what could be achieved by each of the two scaffolds alone. Thus, our results make a strong case for augmenting online DBL by scaffolding measures directed at both the interactional and the content-related level. A limitation of our study, however, may lie in the measurement of web-design knowledge and skills, since our operationalizations merely relied on rather context-free, one-dimensional measures. For example, measuring the acquisition of design knowledge according to different categories such as stages, values, roles, principles, patterns, techniques, and design psychology, as suggested by Hoadley and Cox (2009), might be a valuable alternative to our approach.

References

Armstrong, S. (2003). *Snapshots: Educational insights from the Thornburg Center*. Lake Barrington, IL: Thornburg Center.

Baker, M. (2003). Computer-mediated argumentative interactions for the co-elaboration of scientific notations. In J. Andriessen, M. Baker & D. Suthers (Eds.), *Arguing to Learn: Confronting Cognitions in Computer-Supported Collaborative Learning Environments* (pp. 47-78): Dordrecht, The Netherlands: Kluwer Academic Publishers.

Brem, S. K., & Rips, L. J. (2000). Explanation and evidence in informal argument. *Cognitive Science, 24*(4), 573-604.

Brown, J. (1991). Television critical viewing skills education: *Major media literacy projects in the United States and selected countries*: Hillsdale, NJ: Lawrence Erlbaum.

Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research, 64*(1), 1-35.

Cox, R., & Brna, P. (1995). Supporting the use of external representations in problem solving: the need for flexible learning environments. *Journal for Artificial Intelligence in Education, 6*, 239-302.

Crismond, D., Camp, P. J., Ryan, M., & Kolodner, J. L. (2001). *Design Rules of Thumb - Connecting Science and Design*. Paper presented at the AERA, Seattle, WA, April 2001.

Fischer, F., Bruhn, J., Grasel, C., & Mandl, H. (2002). Fostering collaborative knowledge construction with visualization tools. *Learning and Instruction*, 12(2), 213-232.

Fortus, D., Dershimer, R., Krajcik, J., & Mamlok-Naaman, R. (2004, November). Design-based science and student learning. *Journal of Research in Science Teaching, 41*(10), 1081 – 1110.

Haake, J. M., & Schümmer, T. (2003). Supporting collaborative exercises for distance learning. In B. Wasson, S. Ludvigsen & U. Hoppe (Eds.), *Designing for Change in Networked Learning Environments* (pp. 125-134). Dordrecht: Kluwer.

Harbeck, J., & Sherman, T. (1999). Seven Principles for Designing, Developmentally, Appropriate Web sites for Young Children. *Educational Technology, 39*(4), 39-44.

Hoadley, C., & Cox, C. D. (2009). What is design knowledge and how do we teach it? In C. DiGiano, S. Goldman & M. Chorost (Eds.), *Educating learning technology designers: Guiding and inspiring creators of innovative educational tools* (pp. 19–35). New York: Routledge.

Hobbs, R. (1998). Building citizenship skills through media literacy education. In M. Salvador & P. Sias (Eds.), *The Public Voice in a Democracy at Risk* (pp. 57-76): Westport, CT: Praeger Press.

Hoppe, H. U., Gaßner, K., Mühlenbrock, M., & Tewissen, F. (2000). Distributed visual language environments for cooperation and learning. *Journal Group Decision and Negotiation*, 9(3), 205-220.

Jacobson, M. J., & Levin, J. A. (1995). Conceptual frameworks for network learning environments: constructing personal and shared knowledge spaces. *International Journal of Educational Telecommunications, 1*(4), 367-388.

Jonassen, D. H., Beissner, K., & Yacci, M. (1993). *Structural Knowledge: Techniques for Representing, Conveying, and Acquiring Structural Knowledge*: Hillsdale, NJ: Lawrence Erlbaum Associates.

King, A. (1989). Effects of Self-Questioning Training on College Students' Comprehension of Lectures. *Contemporary Educational Psychology, 14*, 366-381.

Kollar, I., Fischer, F. & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning & Instruction*, 17(6), 708-721.

Kollar, I., Fischer, F., & Hesse, F. W. (2006). Collaboration scripts - a conceptual analysis. *Educational Psychology Review, 18*(2), 159-185.

Kolodner, J. L. (2002). Facilitating the Learning of Design Practices: Lessons Learned from an Inquiry into Science Education. *Journal of Industrial Teacher Education, 39*(3), 1-31.

Piette, J., & Giroux, L. (2001). The theoretical foundations of media education programs. In R. W. Kubey (Ed.), *Media Literacy in the Information Age: Current Perspectives* (pp. 89-134): New Brunswick, NJ: Transaction Publishers.

Powell, G. C. (2001). The ABCs of Online Course Design. *Educational Technology, 41*(4), 43-47.

Romance, N. R., & Vitale, M. R. (2002). Concept mapping as a tool for learning. *College Teaching, 47*(2), 74-79.

Sadler, P. M., Coyle, H. P., & Schwartz, M. (2000). Engineering Competitions in the Middle School Classroom: Key Elements in Developing Effective Design Challenges. *The Journal of the Learning Ssciences*, 9(3), 299–327.

Scott, S. (1995). The technological challenge for curriculum and instruction. *Journalism and Mass Communication Educator, 2*, 30-37.

Shannon, L. (2008). Information and communication technology literacy issues in higher education. *Information Systems Education Journal, 6*(23), 1-13.

Spyridakis, J. H., Wei, C., Barrick, J., Cuddihy, E., & Maust, B. (2005). Internet-based Research: Providing a Foundation for Web Design Guidelines. *IEEE Transactions on Professional Communication, 48*(3), 242 - 260.

Suthers, D. D. (2003). Representational Guidance for Collaborative Learning. In H. U. Hoppe, F. Verdejo & J. Kay (Eds.), *Artificial Intelligence in Education* (pp. 3-10). Amsterdam: IOS Press.

Tabak, I. (2004). Synergy: A complement to emerging patterns of distributed scaffolding. *Journal of the Learning Sciences*, 13(3), 305–335.

Tergan, S.-O., & Keller, T. (Eds.) (2005). *Knowledge and information visualization – Searching for synergies*. Heidelberg: Springer.

Vattam, S. S., & Kolodner, J. L. (2006). *Design-based science learning: Important challenges and how technology can make a difference*. Paper presented at the Paper presented at International Conference of the Learning Sciences.

Wallace, P., & Clariana, R. (2005). Perception versus reality: Determining business students' computer literacy skills and need for instruction in information concepts and technology. *Journal of Information Technology, 4*, 141-150.

Weinberger, A., Stegmann, K., & Fischer, F. (2010). Learning to argue online: Scripted groups surpass individuals (unscripted groups do not). *Computers in Human Behavior*, 26, 506-515.

Weinberger, A., Ertl, B., Fischer, F., & Mandl, H. (2005). Epistemic and social scripts in computer-supported collaborative learning. *Instructional Science*, 33(1), 1-30.

Collaborative Writing: Too Much of a Good Thing? Exploring Engineering Students' Perceptions Using the Repertory Grid

Anindito Aditomo, CoCo Research Centre, Sydney University, and Faculty of Psychology, Surabaya University, aadi4954@uni.sydney.edu.au
Rafael A. Calvo, School of Information Technologies, Sydney University, rafael.calvo@sydney.edu.au
Peter Reimann, CoCo Research Centre, Sydney University, peter.reimann@sydney.edu.au

Abstract: Students' perceptions of technology-supported collaborative writing (CW) and peer reviewing are important because these perceptions affect adoption decisions, initial engagement, and continued reliance on collaboration/peer feedback as a learning resource. This paper describes and demonstrates the utility of the Repertory Grid Technique to probe such perceptions. Combining interviews and written surveys, this study uncovered constructs that engineering students' use to think about CW/peer-reviewing, and interesting patterns of relations among those constructs. For instance, while students perceived activities that involve the construction of personal arguments (such as CW) to be exciting and effective for learning, they also judged such activities to be time-consuming and stressful. This study also found interesting differences in the construct systems of high- and low-performing students.

Introduction

The potential of writing as a learning activity has long been recognized (e.g., Emig, 1977). From a cognitive perspective, writing has the potential to foster deep understanding of the subject matter. This is because writing involves the construction of, and interplay between, conceptual problems (what to say) and rhetorical problems (how to say it) (Bereiter & Scardamalia, 1987). From a socio-cultural perspective, and more specifically through the lens of genre theories, writing assignments provide a context in which students can acquire and practice rhetorical skills and strategies useful to integrate into a discipline-specific discourse community (Gee, 2004).

The importance of writing in higher education is well recognized. Recent nationally commissioned reports from the US and UK, for instance, highlight the central role of writing in graduates' ability to participate effectively in the knowledge economy (Davies, Swinburne, & Williams, 2006; National Commission on Writing in American Schools and Colleges, 2003). Looking at engineering education more specifically, writing is also recognised as a key graduate attribute. And because much of writing in professional practice is done collaboratively (e.g., Ede & Lunsford, 1992 show that 85% of the documents produced in offices and universities have at least two authors), there is a clear need for developing students' collaborative writing skills.

Although there is plenty of evidence for writing as a potent learning activity, there is no guarantee that students will enact their writing assignments productively. One way to enhance the probability that students will use writing activities for deep learning is to embed writing in a social context. A rudimentary, but frequently used form of writing consists of individual or collaborative writing (CW) combined with peer reviewing (Topping, 2005). A major advantage of CW and peer review is that they can be conducted with large numbers of students. Indeed, CW and peer reviewing can be seen as an answer to the specific pedagogical challenge that arises when employing writing in settings with a large number of students, e.g. for undergraduate education: How can guidance (scaffolding) and (formative) feedback on writing be provided, given the teacher:student ratio? This necessitates looking for alternative resources, in the form of self-guided learning, peer feedback, and guidance/feedback that can be provided by computational means.

Our research team have begun to develop an online writing environment with components specifically designed to support the peer review process (Calvo, O'Rourke, Jones, Yacef, & Reimann, 2010), reflection on writing products (Villalon, Kearney, Calvo, & Reimann, 2008), and reflection on writing process (Southavilay, Yacef, & Calvo, 2010). Our inquiry into CW and peer reviewing currently follows two lines. First, we are interested in understanding how students write by examining process data. Second, we are also interested in students' perceptions and subjective experiences of CW and peer reviewing. The study reported here is part of this second line of inquiry. The study aimed to explore engineering students' perceptions of CW and peer reviewing. In addition, the study also examined the utility of a variant of the Repertory Grid Method as an approach to assess the structure of students' perceptions.

The paper begins with a brief summary of prior studies on students' perceptions of writing, along with the theoretical perspective informing those studies. Next, the repertory grid approach to capture students' perceptions is described, followed by the study's data collection methods, analysis, and main findings. We close by discussing the implications for the design and integration of computer supported CW pedagogy and technology into undergraduate engineering education.

Students' Perceptions/Conceptions of Writing

We are interested in students' perceptions of CW and peer reviewing, and in their perceptions of the technologies made available to them, because these perceptions may affect not only how and to what extent students engage with the pedagogy and the technology initially, but also be an important factor affecting the self-guided employment of writing and of peer support as tools for continuous professional development.

One research tradition that provides a theoretical framework that links students' perceptions, learning behavior, and subsequent outcomes is phenomenography (Marton & Booth, 1997). Phenomenography's key insight is that learning is relational: a learning activity is always the product of the relation between the student and the task. Students who see a task differently will enact it in different ways. Phenomenographic studies have identified two basic ways of conceiving a learning task: *cohesive* (as orienting towards transforming personal understanding) and *fragmented* (as orienting towards collecting information or memorizing). Students adopting a cohesive conception typically engage at deeper cognitive levels than those adopting a fragmented conception (Marton & Saljo, 2005).

Several phenomenographic studies have investigated students' views of writing. For instance, based on interviews with history and psychology students, Hounsell (1984) identified different conceptions of essay writing. Some students talked about essays as a presentation of a personal argument or perspective; other talked about essay writing as presenting a string of ideas and facts/data, without an overall organizing argument or perspective. A study by Prosser and Webb (1994) found that students' adopting the more cohesive conception (i.e. essay are argument) also adopted a deeper approach (e.g. making decisions on what needs to be included in the essay based on the overall argument) and produced better quality essays. These associations have also been demonstrated by more recent, larger scale studies (Ellis, Taylor, & Drury, 2007).

While phenomenography has proved to be a useful framework, studies informed by phenomenography have almost exclusively conceptualized experience in terms of the cohesive-fragmented and deep-surface distinctions. We suspect that students' perceptions and experiences of CW and peer reviewing may include other dimensions that are also important to explore. Hence, we need a method to capture students' perceptions without relying too much on a priori or pre-determined dimensions. To this end, we see the Repertory Grid Technique (RGT) as a promising approach. The RGT was developed mainly in psychology based on George Kelly's Personal Construct Theory (Kelly, 1991). Kelly's theory is compatible with phenomenography: both agree that what matters most is not reality as such, but how people see or experience it. Kelly proposed that differences in people's perceptions the world are due to (not necessarily explicit) theories that people hold about respective aspects of the world, and that these theories affect perception through *constructs*, defined as "...a way in which some things are construed as being alike and yet different from others" (Kelly, 1991, p. 74). Constructs develop and change over time as a function of experience, but are relatively stable compared to the speed with which situation and context change. Furthermore, Kelly suggested that each individual possesses a finite number of constructs at any point in time, and that these constructs are related to each other to form a system. The next section describes RGT in more detail.

The Repertory Grid Method

One of our aims in this study is to identify a method that can help capture students' perceptions with the goal of improving upon the design of pedagogy and technology, as well as gauging the effects of differences in perceptions on learning outcomes. This requires a method that works with a small number of students, delivering rich data necessary for informing design decisions, as well as with larger number of students, necessary for generalizing findings. The method we employed to satisfy these demands is based on Repertory Grid Technique (RGT, Fransella, Bell, & Bannister, 2004).

RGT has a long and productive history in social psychology and clinical psychology, but also in areas such as marketing research and information systems research (see Curtis, Wells, Lowry, & Higbee, 2008; Tan & Hunter, 2004 for reviews of applications in the information systems area). A major advantage compared to survey methods is that RGT does not only elicit peoples' ratings but also the dimensions along which these ratings are made. In its canonical format, RGT proceeds by presenting participants with a set of *elements* (the objects of perception, such as social partners, car models, or in our case: teaching/learning methods and technologies) and then eliciting from participants in a systematic manner the constructs they use to distinguish and to group the elements. Importantly, the constructs, or dimensions for comparison, are not given to the participants, but generated by them during grid elicitation.

A typical method to elicit constructs would be an interview where the interviewer presents the participant with sets of three elements (e.g., three learning technologies) and then asks the participant to select two elements that are similar, and distinct from the third, and then to label the dimension that forms the basis for this triangulation (e.g., interactivity with the poles Low and High). Repeating this triangulation for all combinations of elements yields a multidimensional space in which each element occupies a position. Over this data representation, one can for instance calculate similarity measures that can be further subjected to qualitative analysis, cluster analysis, or dimension reduction methods (e.g., principal component analysis; multidimensional scaling). Such data can be analyzed on the individual level as well as pooled over participants. For more descriptions of the various elicitation and analysis methods, as well as software tools, the reader is referred to tutorials such as Fransella et al. (2004).

A major practical drawback of RGT is the time it takes to conduct the construct elicitation, in particular if the elicitation is done by interview. Since (engineering) students have little time, and we did not have the means to pay our research participants, we employed a combination of RGT and standard survey methodology. A small number of volunteering students were interviewed with the goal to elicit their construct system, and the main constructs were then used as the basis for a survey that was distributed amongst a larger number of students. While this has the disadvantage of potentially glossing over interesting individual construct systems, it has the advantages of being much more rapidly applied and of reducing interviewer bias.

Research Questions and Methods

This study addressed two main questions: How do engineering students' perceive CW and peer reviewing? And are there any relations between students' perceptions and their writing performance?

Participants and Course Context

Participants in this study were 3rd-year engineering students enrolled in a course titled E-Business Analyses and Design in Semester 1, 2010. Nine students (4 males, age 20 to 23) from the course volunteered to participate in the interviews. Thirty-nine students (31 male) volunteered to participate in the survey. (This sample represented 73.6% of the whole class.) The course focused on "aspects of analysis, project specification, design, and prototype that lead up to the actual build of a website or application."

Of particular interest for this study was the writing assignment, which required students to work in pairs and write an e-business proposal. While this was a group task, there was a section of the proposal that each student had to write individually to obtain individual marks. After the submission of the proposal, each student had to read and write a review of another group's proposal (the peer review was also graded). Students were then asked to revise their submitted proposal based on the peer review and tutor feedback. The CW and peer reviewing were performed in an online environment (iWrite).

Online Writing Environment

Our on-line writing environment ('iWrite') provides students with tutorials on writing for different disciplines, assignment information, and tools for writing, improving their writing, reviewing and submitting their assignments (Calvo, O'Rourke, Jones, Yacef, & Reimann, 2011). To instructors, it provides a complete solution for scaffolding the write-review-feedback cycle of a writing activity. Specialist writing instruction content is built into the technology, and management tasks which would be highly time-consuming for large classes are handled automatically. In the course investigated here, iWrite was used to support an assignment that combines CW and peer reviewing.

Data Collection

Six interviews were conducted to elicit the constructs students used to think about learning activities. Three of the six interviews were conducted with pairs of students. Each interview took 20-30 minutes. In the interviews, students were shown triads of learning activities and were asked: "Here are three learning activities. Choose two among these three, and think about how they are similar to each other, but different from third." This procedure was first demonstrated using an example triad: taxi/private car/public bus. The learning activity triads were: (a) reading textbooks/reading articles/attending lectures, (b) Peer reviewing/individual oral presentations/group oral presentations, (c) attending lectures/tutorials/peer reviewing, (d) writing individually/writing collaboratively/reading textbooks. For each triad, the interviewer wrote a phrase or word to represent the student's answer as his/her construct and asked whether the student agreed with the phrase/word.

Each of the interviews produced 5 to 11 constructs (median: 7). While some constructs were commonly present in 2 or 3 of the interviews, no constructs appeared in all 6 interviews. More in-depth analysis can be done on the interview data to explore the interviewees' individual construct system. This paper, however, focuses on the pattern of perception across the whole class sample. For the purpose of this paper, we report that the interviews elicited a total of 31 constructs. Because a pilot study indicated that we could only include approx. 10 constructs (given the time constraints), we had to select from the pool of 31 constructs. We did so by first examining whether the 31 constructs could be grouped based on their similarity in meaning. This resulted in 15 groups of constructs. From this, for the survey, we selected 11 constructs that were mentioned by more than 1 student. These 11 constructs are listed in Table 1 in the next section, and the 4 excluded were: "interactive/non-interactive", "provides clues about important materials/not", "requires communication skills/not", and "depends on one's own motivation/supervised".

The survey was distributed on paper and online. The questionnaire asked students to rate five learning activities which they experienced in the course, plus what students' consider to be an "ideal" learning activity, on a scale from 1 to 5 in terms of the 11 constructs elicited from the interviews.

Findings

Interrelations between the Constructs

To explore the interrelations between constructs students used in thinking about learning activities, we first performed an exploratory factor analysis (Oblimin rotation with Kaiser Normalization was used). The factor analysis suggested a 4-factor solution which accounts for 67.26% of the total variance. Sampling adequacy (KMO) was measured at 0.644, which is acceptable, while Bartlett's sphericity test was significant (chi-square=587.8, p< .0001). The factor analysis results are shown in Table 1.

Table 1: Factor analysis results (loadings of less than .3 are not shown).

#	Constructs	Factor loadings			
		1	2	3	4
1	Requires construction of own argument (1) vs. does not (5)	.789			
2	Requires independent research (1) vs. does not (5)	.763			
3	Involves peer discussion (1) vs. no peer discussion (5)	.578		.409	
4	Boring (1) vs. exiting (5)	-.547	.510		-.388
5	Effortful (1) vs. effortless (5)		.782		
6	Time consuming (1) vs. not time consuming (5)		.754		
7	Stressful (1) vs. relaxing (5)		.738		-.391
8	Can be done anywhere (1) vs. only in certain places (5)			-.887	
9	Can be done anytime (1) vs. only in certain times (5)			-.869	
10	Structured (1) vs. unstructured (5)				.807
11	Effective (1) vs. not effective (5)	.318	-.355		.726

The first factor may be interpreted as "Student-centeredness", with more student-centered learning activities requiring students to perform their own research, construct their own argument, and involve peer collaboration. The second factor may be interpreted as "Demand", with more demanding activities being effortful, time consuming, stressful, and boring. The third factor was interpreted as "Flexibility", with more flexible activities being constrained by neither time nor location. The fourth factor is more difficult to interpret, as it is composed of "Structuredness" and "Effectiveness for Learning". This implies that that students tend to perceive structured activities as being more effective than unstructured ones.

The cross-loading of the "Boring/exiting" construct is worth noting. On the one hand, this construct loads with the "Student-centeredness" factor, indicating that students tend to perceive student-centered activities as exiting. "Boring/exiting" also loads with the "Structure and Effectiveness" factor, showing that activities perceived to be structured are also perceived to be boring. On the other hand, "Boring/exiting" loads with the "Demand" factor, implying that activities perceived as demanding are also perceived as boring. This pattern highlights the challenge of engaging engineering students: to be perceived as exiting, a learning activity needs to be unstructured and student-centered, but at the same time relaxing and requiring little effort and time.

The cross-loading of the "Effectiveness" construct is also interesting to note, in particular because it follows the opposite pattern to the "Boring/exiting" construct. While "Effectiveness" loads most strongly with "Structuredness", it also loads with the "Student-centeredness" and "Demand" factors. The directions of these loadings indicate that activities which are more student-centered and demanding (time/effort) were also seen to be more effective for learning.

Students' Perceptions of Collaborative Writing and Peer Reviewing

How do engineering students perceive CW and peer reviewing, *vis a vis* other learning activities? In addressing this question, we used the factor analysis results above to simplify the 11 original constructs by combining those which loads strongly (above .7) and exclusively onto a single factor. This resulted in the 7 constructs: (1) *Student-centeredness* (combination of "independent research" and "construction of argument"), (2) *Effectiveness*, (3) *Structuredness*, (4) *Demand* (combination of "time", "effort", and "stress"), (5) *Flexibility* (combination of "anywhere" and "anytime"), (6) *Excitingness*, and (7) *Peer discussion*. (Note: to further aid interpretation, some of the scores have been reversed such that higher scores reflect stronger perception in that construct; possible score ranges from 1 to 5 for all constructs).

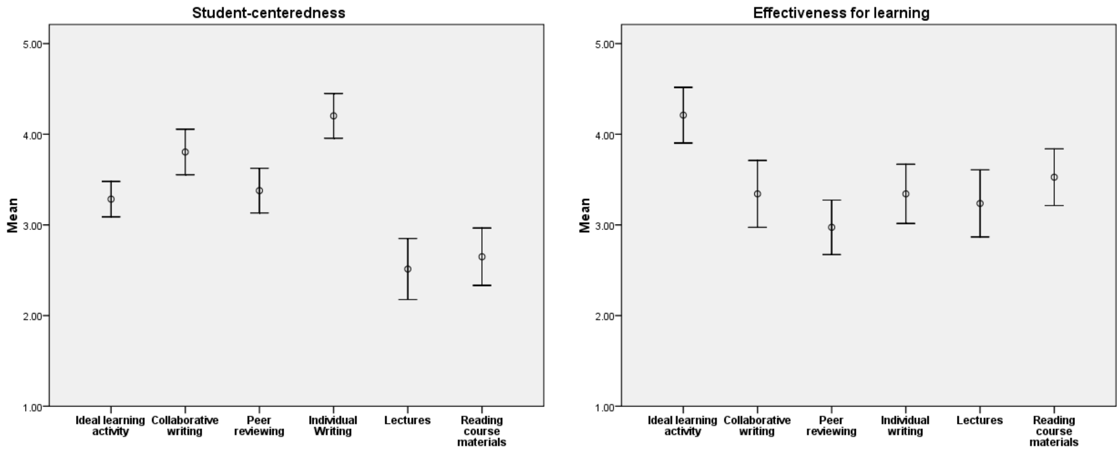


Figure 1. Perceived Student-Centeredness and Effectiveness; Error Bars Represent 95% Confidence Intervals.

As Figure 1 (left panel) shows, CW was seen to be the second most student-centered activity (after individual writing), while attending lectures and reading course materials the least. Interestingly, both collaborative and individual writing were seen as more student-centered than an ideal learning activity. In other words, writing was seen to involve *too much* construction of personal argument and independent research. Peer reviewing, on the other hand, was considered to be ideal in terms of student-centeredness.

In terms of learning benefits (Fig. 1, right panel), CW, individual writing, and peer reviewing were all seen as no more effective than attending lectures. All learning activities fell short of students' expectations in this regard. Peer reviewing, in particular, was seen to have the least benefit for the students' learning; it is seen as slightly less effective than CW (Wilcoxon $z = -2$, $p = .045$), individual writing (Wilcoxon $z = -1.72$, $p = .085$), and reading (Wilcoxon $z = -2.331$, $p = .02$).

There seems to be relatively little variation in the perceived level of structure across the five learning activities (Figure 2, left panel). CW and peer reviewing were seen to be the least structured activities, although only slightly less structured than ideal (the difference between ideal and peer reviewing was statistically significant, $z = -2.242$, $p = .025$).

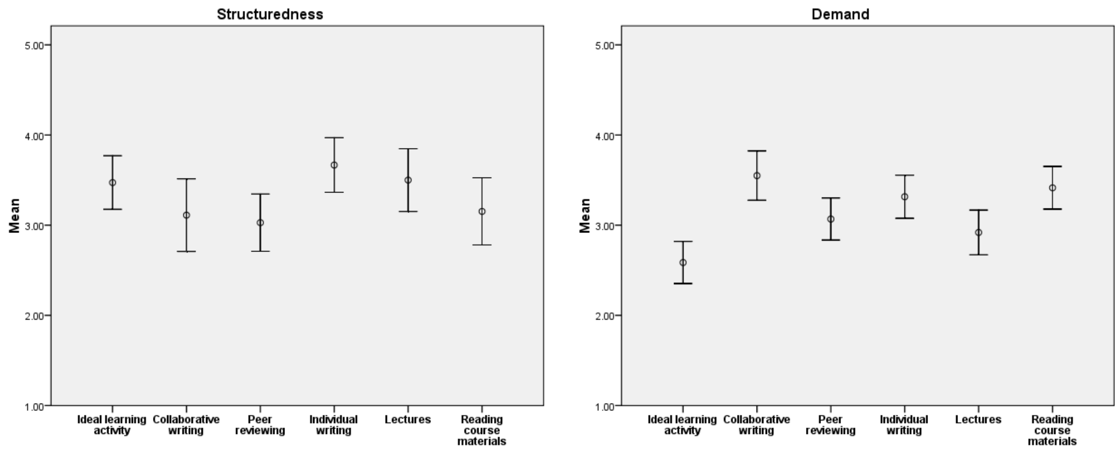


Figure 2. Perceived Structuredness and Demandingness; Error Bars Represent 95% Confidence Intervals.

In terms of perceived demand (Figure 2, right panel), individual writing, peer reviewing, and (in particular) CW were all judged to demand more time/effort and more stressful than an ideal learning activity. Peer reviewing seems to pose slightly less demand than CW. While it is quite easy to see why the writing-related activities were seen as demanding, it is worth noting that the lectures and particularly reading materials were also deemed too demanding. Hence, designing a learning activity that meets students' expectations in this regard may be difficult.

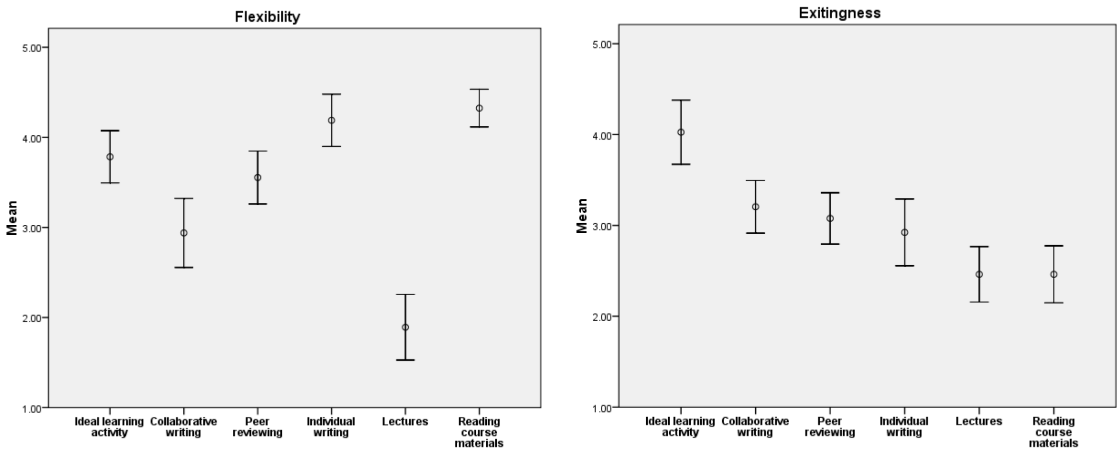


Figure 3. Perceived Flexibility and Excitingness; Error Bars Represent 95% Confidence Intervals.

With regards to flexibility (Figure 3, left panel), attending lectures was seen to be the least flexible, while individual writing and reading were the most. Interestingly, CW was considered to be less flexible than ideal (although still more flexible than attending lectures), despite the availability of the online writing environment. Peer reviewing was seen to be more flexible than CW, but still slightly less flexible than ideal. With regards to level of excitement (Figure 3, right panel), none of the five learning activities met students' expectations. However, CW, peer reviewing, and individual writing were all perceived to be more exciting than attending lectures and reading course materials. CW was unsurprisingly seen to involve the most peer discussion. This, however, was more than what students would have liked. On the other hand, peer reviewing and individual writing, along with the other learning activities, were all seen to involve too little peer discussion.

Relations between Students' Perceptions and Writing Performance

Do students' perceptions of CW and peer reviewing predict their writing marks? Simple non-parametric correlation tests were performed to examine this question. Results indicated that students who obtained better writing marks also tended to perceive: CW to be more effective for learning ($r = .348$; $p= .032$) and more exciting ($r = .298$, $p = .065$); and peer reviewing to be less structured ($r = - .342$, $p = .036$), involve less peer discussion ($r = - .376$, $p = .02$), and more exciting ($r = .287$, $p = .07$). These significant correlations are mostly in line with expectations: students who see CW and peer reviewing as exciting and/or beneficial for learning would likely adopt more productive approaches in their writing. The other two correlations (between higher marks with perceptions of less structure and peer discussion in peer reviewing) are less obvious and need further exploring.

In addition to the correlation analysis, we also compared the construct system of high and low performing students. This analysis utilizes the unique structure of repertory grid data, which allows the application of principle components analysis on the constructs to produce a 2-dimensional map whose axes represent the first two extracted factors, onto which each learning activity can be plotted. This kind of analysis has the advantage of depicting the structure of students' whole construct system, as opposed to looking at single constructs separately. For this analysis, the high performing group comprised of six students who scored 8 or more (out of 10), while the low performing group comprised of five students who scored 4 or less in their CW assignment. The average construct ratings were calculated as a basis of the principle components analysis, which was performed using Chris Evans' program (http://www.psyctc.org/grids/). The results (Figure 5) showed some striking differences between high and low performing students.

In the construct system of high performers (Figure 5, left panel), CW is placed quite close to the ideal learning activity. More specifically, CW is considered to be exiting, effective, student-centered, involves a lot of peer discussion, unstructured, and simultaneously posing low demand. In contrast, in the construct system of low performers (Figure 5, right panel), CW is placed far from the ideal learning activity and is perceived to be ineffective, unstructured, not exiting, and not student-centered.

Another difference between high and low performers is related to the construct of "structuredness": in the high-performers' construct system, "structuredness" is associated with level of demand, but not necessarily with student-centeredness, effectiveness, nor excitingness. The opposite is true in the construct system of low performers: "structuredness" is strongly associated with student-centeredness, effectiveness, and excitingness. In other words, low performers (but not for high performers) seem to expect more structure to be able to see the learning and motivational value of an activity.

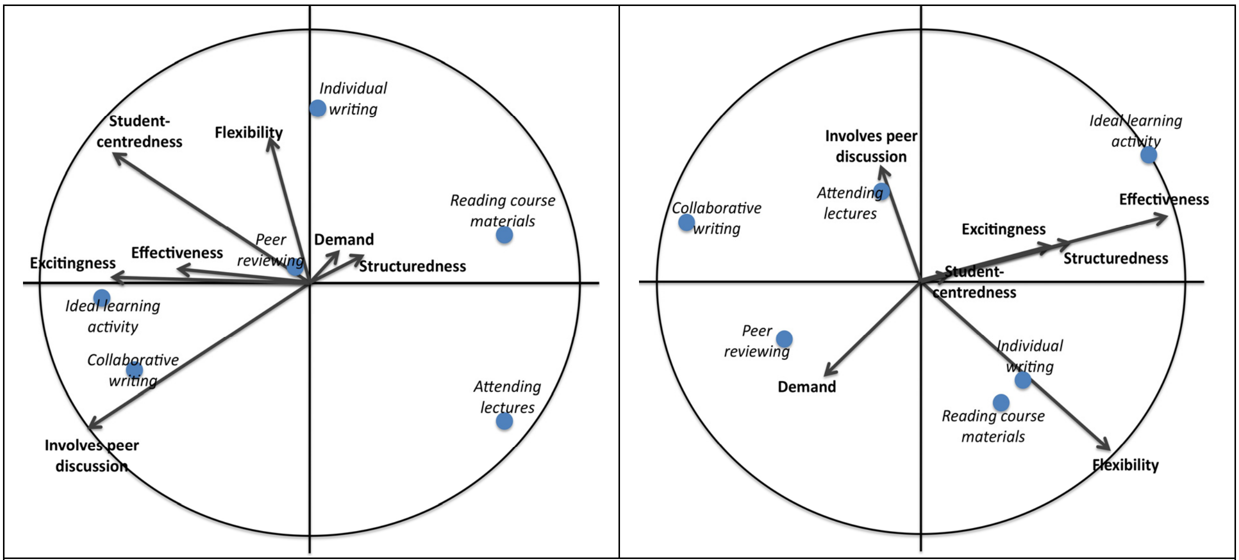


Figure 5. Construct System of High Performing (Left Panel) and Low Performing (Right Panel) Students.

Discussion and Conclusions

This study aimed, firstly, at revealing students' perception of (technology-supported) peer review pedagogy, an arguably rudimentary but frequently employed method for collaborative learning in undergraduate education. Secondly, we were interested in relations between students' perceptions and their success on writing assignments.

We found that writing, particularly when done in groups, was perceived to be quite student-centered. This indicates students' awareness that writing requires them to do some independent research and construct their own argument. CW and peer reviewing were seen to be moderately exciting, more so than attending lectures and reading. However, on average, writing and peer reviewing were seen to be only moderately effective for learning, not more effective compared to attending lectures and reading. In addition, while individual writing was (as predicted) considered to be flexible, CW was considered to be only moderately flexible. This indicates that in performing their writing assignment, many of these students still see the need to have face-to-face meetings, despite the possibility of online, asynchronous collaboration afforded by the technology.

Comparisons with students' imagined ideal learning activity provided some more insights. CW was seen to be ideal only in terms of structuredness. In terms of the other constructs, however, CW did not meet students' expectations: it was seen to be *too* student-centered and demanding, and at the same time not effective and flexible enough. Peer reviewing was as more ideal in terms of student-centeredness, structuredness, and flexibility, but still less effective and slightly more demanding than ideal.

These findings highlight some of the challenges of introducing CW to engineering students. If CW is seen as too stressful and demanding too much time/effort, while at the same time providing little learning benefits, then it would be difficult to persuade engineering students to adopt a deep approach to their writing assignments. On the positive side, these students still considered writing and peer reviewing to be more exciting and student-centered than attending lectures and reading course materials. Hence, the students were aware of some of the positive aspects of CW and peer reviewing. It is this dimension of students' perception that lecturers can tap into when trying to motivate students to adopt productive approaches to CW/peer reviewing.

This study also demonstrates that the repertory grid technique (RGT) can be fruitfully applied to explore the multi-faceted nature of students' perceptions. The specific approach we adopted – using constructs elicited through interviews in a follow-up survey – proved both useful and practically viable. Although it may gloss over individual students' unique construct systems, this approach provided insights into students' complex view of CW and peer reviewing. For instance, from the findings we can point out specific aspects of CW that students appreciate and other aspects which they don't. We were also able to demonstrate not only *whether*, but *how* high performers' construct systems differ from the low performers'.

Furthermore, the approach employed here yielded insights that enrich the traditional findings of phenomenographic studies. Whereas phenomenography focuses on cohesive and fragmented conceptions of learning, our study suggests that affective dimensions (e.g. exitingness) and pragmatic effort calculation (e.g. demandingness) are also important parts of students' experience of CW and peer reviewing. When we consider a richer set of dimensions of students' perceptions, it is difficult to dismiss the engineering students in this sample as mere surface learners: While most saw CW and peer reviewing as demanding tasks, they were also aware that those tasks require time, effort, independent work, and the construction of personal arguments.

References

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Mahwah, NJ: Lawrence Erlbaum Publishers.

Calvo, R. A., O'Rourke, S. T., Jones, J., Yacef, K., & Reimann, P. (2010). Collaborative writing support tools on the cloud (accepted for publication. *IEEE Transactions on Learning Technologies*.

Calvo, R. A., O'Rourke, S. T., Jones, J., Yacef, K., & Reimann, P. (2011). Collaborative writing support tools on the cloud. *IEEE Transactions on Learning Technologies, 4*(1), 88-97.

Curtis, A. M., Wells, T. M., Lowry, P. B., & Higbee, T. (2008). An overview and tutorial of the repertory grid technique in information systems research. *The Communications of the Association for Information Systems, 23*, 38-62.

Davies, S., Swinburne, D., & Williams, G. (Eds.). (2006). *Writing Matters: The Royal Literary Fund Report on Student Writing in Higher Education*. London: The Royal Literary Fund.

Ede, L. S., & Lunsford, A. A. (1992). *Singular texts/plural authors: Perspectives on collaborative writing*: Southern Illinois Univ Pr.

Ellis, R. A., Taylor, C. E., & Drury, H. (2007). Learning science through writing: associations with prior conceptions of writing and perceptions of a writing progam. *Higher Education Research & Development, 26*(3), 297-311.

Emig, J. (1977). Writing as a mode of learning. *College Composition and Communication, 28*(2), 122-128.

Fransella, F., Bell, R., & Bannister, D. (2004). *A manual for repertory grid technique* (2nd ed.). West Sussex: Wiley.

Gee, J. P. (2004). Language in the science classroom: academic social languages as the heart of school -based literacy In E. W. Saul (Ed.), *Crossting Borders in Literacy and Science Instruction: Perspectives in Theory and Practice* (pp. 13-32). Newark, DE: International Reading Association.

Hounsell, D. (1984). Essay planning and essay writing. *Higher Education Research & Development, 3*(1), 13-31.

Kelly, G. A. (1991). *The psychology of personal constructs (Original work published 1955)*. London: Routledge.

Marton, F., & Booth, S. (1997). *Learning and Awareness*. New Jersey: Mahwah.

Marton, F., & Saljo, R. (2005). Approaches to learning. In F. Marton, D. Hounsell & N. Entwistle (Eds.), *The Experience of Learning: Implications for Teaching and Studying in Higher Education* (3rd (Internet) ed., pp. 39-58). Edinburgh: University of Edinburgh.

National Commission on Writing in American Schools and Colleges. (2003). *The Neglected "R": The Need for a Writing Revolution*. New York: College Entrance Examination Board.

Prosser, M., & Webb, C. (1994). Relating the process of undergraduate essay writing to the finished product. *Studies in Higher Education, 19*(2), 125-138.

Southavilay, V., Yacef, K., & Calvo, R. A. (2010). *Process mining to support students' collaborative writing.* Paper presented at the Educational Data Mining, Pittsburgh, PA, USA.

Tan, F. B., & Hunter, G. M. (2004). Cognitive research in information systems using the Repertory Grid technique. In M. E. Whitman & A. B. Woszcsynski (Eds.), *The handbook of information systems research* (pp. 261-290). Hershey, PA: Idea Group Inc.

Topping, K. J. (2005). Trends in Peer Learning. *Educational Psychology, 25*(6).

Villalon, J., Kearney, P., Calvo, R. A., & Reimann, P. (2008). *Glosser: Enhanced feedback for student writing tasks*. Paper presented at the 8th IEEE International Conference on Advance Learning Technologies (ICALT) (Santadar, Spain, July 1-5, 2008; accepted March 2008).

Acknowledgements

This project was funded by an Australian Research Council Discovery Project grant (DP0986873). We thank the students who have generously given their time to participate in this study.

The Interactional Organisation of Location-based Gaming

Jo Dugstad Wake, Frode Guribye, Barbara Wasson, Department of Information Science and Media Studies, University of Bergen. Fosswinckelsgate 6, 5007 Bergen.
Email: jo.wake@infomedia.uib.no, frode.guribye@infomedia.uib.no, barbara.wasson@infomedia.uib.no

Abstract: This paper describes a study aimed at gaining insight into the interactional accomplishment of collaborative game play. Mobile, location-based games are commonly believed to hold great potential for creating novel and immersive learning experiences, but little have been done to unravel the practical details of how such games are actually played. To address this issue, a mobile, location-based game for teaching and learning history have been designed, deployed and the participants' activities were video taped. The focus of the study is how the participants make use of the resources available in the game space and how these resources, including the historical narrative, feature in the participants' practical accomplishment of the game.

Introduction

This paper describes a study aimed at gaining further insight into the interactional organisation and practical accomplishment of collaborative game play. A location-based game for learning history has been designed and deployed and we have video taped students playing the game. The analysis focuses on how the participants used the resources available to them in the game space: the game itself on the GPS-enabled mobile phone, and the urban environment and physical surroundings. Playing the game involves moving around in a city landscape and focus is also set on how the participants make these resources available for each other and how they engage with the material presented in the game instructions. This includes how the participants relate to the historical narrative presented in the game, in and through their interaction, and how this is seen in relation to the historical aspects of actual locations and surroundings. Much of the game play is constituted of mundane, everyday activities such as way finding and orientation in the urban environment, and attention is also set on how these activities are interactionally accomplished and how the resources and their knowledge of the local geography feature in the activity.

The text has the following structure. The first section is dedicated to a short review of relevant studies, such as studies entailing the use of mobile technology and games within the field of CSCL, and video-based research. Then we describe SILO, our technological architecture for creating mobile games, and the mobile game, Premierløitnant Bielke (PB). Next we present our study, including the research focus and methodological approach, and the analysis. Finally we present and discuss our findings.

Mobile, Location-based Computer Games and Learning

The use of mobile technology to support collaborative learning has been discussed within the field of CSCL for almost a decade (see e.g., Roschelle & Pea 2002; Roschelle, Rosas & Nussbaum, 2005). Tools have been developed and studied, both to support collaboration in the classroom (Chang, Wang, Chen & Liang, 2009; White, 2006), and to provide support when moving into the field (Lyons, 2009; Yatani, Sugimoto & Kusunoki, 2004; Tan, Liu & Chang, 2007).

Computer games have also become increasingly used and studied for their educational potential (e.g., McFarlane, Sparrowhawk & Heald, 2002; Kirriemuir & McFarlane, 2004; Shute, Rieber, & Van Eck, in press), also within the field of CSCL (e.g., Ke, 2007; Rosenbaum, Klopfer, Broughner & Rosencheck, 2007; Satwitcz & Stevens, 2007; Klopfer, Perry, Squire, Jan & Steinkuehler, 2005). Bennerstedt and Linderoth (2009), for example, have studied gamers' practices within multiplayer online role-playing games to discover how they are dependent on generic social skills such as use of language. They approach the game of World of Worldcraft as a knowledge domain, and through the use of interaction analysis they show how the participants in the game "make visible what they see as relevant knowledge objects" (Bennerstedt & Linderoth, 2009, p. 410).

The terms ubiquitous and pervasive computing shares a high degree of similarity (Dourish, 2004), and have been used to refer to technology and computer *use* that is not limited to taking place in front of a screen connected to a stationary computer. Technologically, pervasive computing usually relies on devices for sensing and positioning the user, such as GPS or Mobile network base stations, and devices for facilitating an information exchange dependent on the user's location, such as RFID tags, 2D barcodes or Bluetooth.

Mobile, location-based games, refers to a subcategory of ubiquitous and pervasive computing. In such games the physical and cultural surroundings, for example an urban area, are made an integral part of the game space and the location of the gamers is a key aspect of the game-playing activity. The games add a digital layer to these environments, creating a game space where players can explore familiar urban spaces in a new context. Thus, location-based, mobile games represent a new kind of game compared to both traditional console-games and physical games; one that has potential for providing novel learning environments, for example by facilitating the embedding of (abstract) concepts in the contexts of their actual use (Kurti, Spikol, Milrad, Svensson & Petterson, 2007; Kurti, Milrad & Spikol, 2007).

Little have been done, however, to unravel the practical details of how such games are actually played, and subjected gaming activity in mobile, location-based games to detailed empirical analysis at the level of interaction. Such analyses have the potential to reveal the interactional organization of the game-play and how different aspects of a given game are made relevant *in situ.* As mentioned above, the analysis presented in this paper aims at offering this kind of insight into the practical accomplishment and interactional organisation of gaming. There is a body of research inspired by ethnomethodology and conversation analysis relevant for this study, especially studies where the use of technological resources features as a central component in the analysis (e.g., Heath & Luff, 2000; Suchman, 2007). A key issue in these studies is the use of video recordings in the analysis (Heath, Hindmarsh & Luff, 2010). Recently, the use of video-based research has gained a certain momentum in CSCL and the learning sciences (for a recent overview see Derry, et al., 2010). As Koschmann, Stahl and Zemel (2006), point out, a key analytical commitment in such studies is "to discover within the recorded materials what the members are actually accomplishing (...) and are making relevant (…) through their interaction" (Koschmann, Stahl & Zemel, 2006, p. 7).

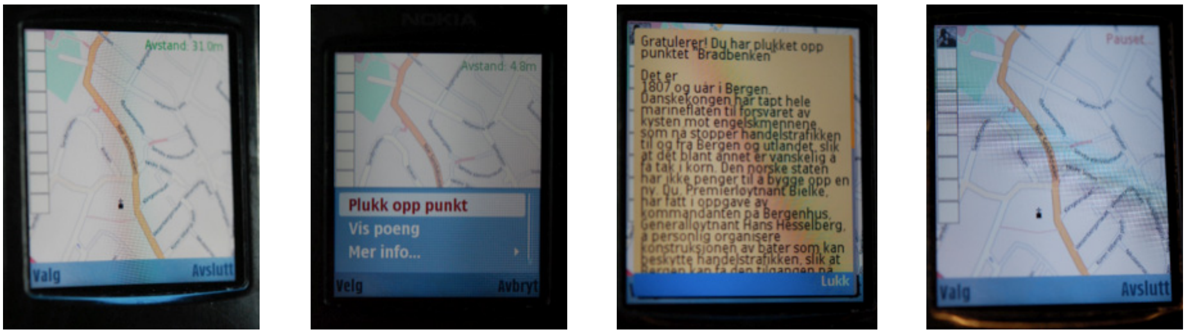
Goodwin (1994), in doing a detailed study of the field work of archaeologists, shows how they make certain aspects of their physical surroundings relevant and visible to each other in interaction, and in this way highlight objects as relevant in the given context. While Goodwin's study is focused on activities that take place in a relatively limited area, his insights are still relevant for understanding how participants, through their interaction as they move through an urban area, give relevance to different aspects of their surroundings and thereby how they perceive these surroundings. Recently, Brown and colleagues (Brown & Chalmers, 2003; Brown & Laurier, 2005; Brown, et al., 2005) carried out a series of ethnomethodological studies of mobile technologies that ranged from simple paper maps to hand held tablet computers. These studies highlight how the participants organised their way-finding and navigation, how they oriented themselves in their geographical surroundings and how they made use of the different available resources. This was all studied at a detailed level of interaction and shows how such activities are practically accomplished and how the use of navigational resources such as a map features in the organisation of these activities. All the above-mentioned studies have informed the analysis presented below. First a description of the game design is presented.

SILO and Premierløitnant Bielke

Premierløitnant Bielke (PB) is a mobile, location-based game for teaching and learning history. It builds on several technological resources, and was created using the SILO system. This section briefly describes SILO and PB (see Wake & Baggetun (2009) for a more detailed account of both).

SILO is a web-based system for creating mobile, location-based games. It permits a game creator to construct a storyline as a set of missions, and attach the different missions to different locations, by clicking on a map, displayed on the screen. Aside from the storyline, the game creator can add icons and set limitations on time, and configure user data, in addition to a maximum of three hints on how to find each location. When all the parameters of a game are specified, they are put in a zip-file that can be transfered to the phone. The mobile application then unwraps the game parameter file, and creates a game. SILO is designed to be easy to use, making it possible for non-technically oriented persons to create games, for example teachers who have an idea for a location-based game for their students.

When the game is being played the application displays a mission (i.e., a description of the next location from the storyline), a marker indicating the participant's current location and the number of meters to the next location, a scrollable map (see Figure 1a), and a track displaying the history of their movement, a set of icons indicating progress in the game, and their current score. While the game is being played the application is constantly calculating the distance to the next location. This is displayed in red until the players near the location – when it turns green and they can pick-up the location (see Figure 1b). They are then congratulated and offered a text that describes the next location (see Figure 1c). The game then pauses, to allow the group to think about what to do next, and an icon signalling that they have picked up the previous location is displayed (see Figure 1d). The game is over when the last mission is solved (i.e., the last location is found). The application continually saves game data so that the game can be restarted and continued from where a game was left (e.g., if a participant accidentally switches off the phone, the participants can restart the game, and continue from where they left off).



(a) (b) (c) (d)

Figure 1. The Mobile Interface of a Game Created by SILO.

The PB Storyline: Bergen during the Napoleonic Wars

The main idea behind the game PB is to use both the concrete geographical surroundings relevant to a historical setting, and a storyline about the same setting in a game, creating an *ad hoc* museum for the purpose of teaching and learning (Wake & Baggetun, 2009). Thus, a novel and immersive experience of history is provided, in a way not possible in the classroom. Furthermore, it is envisaged that the presence of actual buildings and sites, relevant to the history in question, will facilitate reflection over the same history. The game PB is about Bergen, Norway during the Napoleonic Wars. This was a period of distress in Bergen, as the war with Great Britain led the British to block trade routes via the sea. Due to the resulting scarcity of food and supplies, the citizens of Bergen needed to amend the situation, and decided to build small, rowed gunboats to defend merchant ships against the larger British frigates. The game is about the building of these gunboats, and the players are to virtually collect their different parts and accessories, by visiting the actual sites where boat parts and accessories were being produced in Bergen, in the early 1800's. In this way, we highlight the different trades and vocations that were important to the marine trade hub that Bergen was during this period.

PB is designed to be played by groups competing to first complete the set of missions. The groups navigate the area of Sandviken, Bergen in search of locations described in the set of missions. Once they have received a mission, they orient themselves and wander to the location with the aid of the available map, location indicator, distance to next location, up to three hints, and the physical surroundings. When they have found the location, they pick it up and receive the next mission. The group can then reorient and make decisions about where they need to go to find the next location. The game ends when they pick up the last location. The group with the lowest score, consisting of a sum of time and number of hints used, wins.

Research Design and Methods

The gathering of video data and their analysis reported in this text is a continuation of a series of empirical studies of the game of PB. The previous experiences with deploying scenarios of the game, has also been reported on elsewhere with focus on usability (Wake & Baggetun, 2009) and on integration with a web-based publishing tool for integration of the gaming activity with classroom learning activities (Baggetun & Wake, in preparation). The overall research design is based on creating learning scenarios inspired by designed based research (Barab & Squire, 2004), where the deployment of the game is studied and this informs the redesign of the game and the scenario.

In order to study the practical accomplishment and interactional organisation of the game playing, we organised a field trial where the main body of participants were master students in a course in CSCW at our department. Twelve participants were divided into five groups of two or three, and filmed while playing PB.

Nine participants were master students at our department. Two of the participants were acquaintances and the last was a PhD candidate from our department. For the master students, the gaming activity followed a lecture on technologies for mobile support of work, and ubiquitous and pervasive computing in general, in addition to a short introduction to the game. The technological structure the game builds on, and the historical background for the game was presented. (For the remaining participants a quick introduction was given on site as part of the briefing.)

The empirical material consists of seven hours and 15 minutes of video recordings. The average duration of the recordings is 1h 26mins, the shortest 1h 17mins and the longest 1h 46mins. The corpus of video data has been logged and transcribed. This material formed the basis of our analysis.

The first two groups were filmed simultaneously with two different video camera setups, one with and one without an extra wireless microphone attached to one of the participants. The remainder of the groups were filmed with the additional wireless microphone attached. Filming location-based game players in an urban milieu represents a set of practical challenges compared to filming classroom activity, largely in the sense that the potential influence of environmental factors on the quality of the video is higher. One is factor is the quality of sound, as noise from, for example traffic, construction work, treading on snow and bystanders' chatter made reconstruction of speech difficult when an extra microphone was not attached to the participants. The recordings with the wireless microphone, however, are of good quality and were relatively easy to transcribe. Another factor is that the choice of camera angle is challenging. When the groups were moving around, it was difficult to capture footage from another angle than behind the group, as the groups' movement was volatile and discrete. This for example made it more difficult to capture what they were looking on at the screen, and their facial expressions. Still, when they stopped to find the locations, we were able to get good footage of most of these situations. We were able to reconstruct what they were looking at on the screen when they were talking about it (from the audio track), however, and also to some degree from the groups' current location and activity.

Analysis of Game Playing

The analysis of the empirical material revealed an emerging structure in the groups' game activity. In this section we present the structure of the gaming activity. The overall activity structure consists of four phases: 1) Briefing, 2) Search and orientation, 3) Arriving at a location, 4) Receiving instructions. There is iteration over the last three phases, until the final location is reached. Each of these is described below.

Briefing. At the beginning of the game, the participants in each group are given a briefing on the aims of the game and how it works by the game facilitator. This includes showing the directions to the first location.

Search and orientation. During the search and orientation phase, the group moves towards a location. Having received and read the instructions, the group carries out an initial orientation and starts walking. When they needed to adjust their direction, they do so in several ways. One way is that the person carrying the phone reads out loud the distance to the next location. A second way is when a group member verbally suggests that the group is moving in the wrong direction. A third way is markedly changing direction, sometimes together with a comment of some sort.

Arriving at a location. In this phase, the groups establish among themselves whether they have arrived at the intended location. Arriving at the location could mean that the GPS numbers turned green indicating that they are within the zone of the location allowing them to "pick up the spot" to proceed in the game. In addition, they have to consider whether they can identify the actual physical location (the building, object or area). The completion of this phase, in terms of picking up the spot, varied. Either they identify the physical location at a distance and then zone in on the location and pick up the spot, or they enter GPS zone, and pick up the point without identifying the actual physical location. This varied with the groups and the different locations in question. Picking up the spot concluded the phase of zoning in on the location.

Receiving instructions. Instructions on how to find a new location are received on the phone, as the starting point for the search for the next location. The instructions are shared in one of two ways. Either the person carrying the phone reads the instructions out loud to the rest of the group, or they position themselves in a way where each of them can each read the instructions on the display simultaneously.

The groups iterated through the last three phases, searching for and orienting towards a location, zoning in on it and receiving new instructions, until they arrive at the final location where they receive a final message containing historical information that concludes the game.

Interaction in the Game

An episode from one of the groups' game play (a group of three) has been selected to illustrate how the participants organised their gaming activity in the different phases and how they make decisions concerning their involvement in the game, how to play it, and how this is organised in and through interaction. (Excerpt 1, translated to English from Norwegian, and anonymised.) The episode is one minute and nine seconds long, and starts where the group is in the search and orientation phase after having picked up five of the total nine spots in the game. They are trying to find the sixth spot and are walking down a road (see also Figure 2, images used with permission of the participants).

Excerpt 1:

```
1   Anders:  No, it is fort- forty se::ven (0.2) and now it is     Walks
             forty three and the next is forty one (2.1) I'm
             sure it's the building that burned down, don't you
             think? Hehe
2   Bård:    (2.3) Recently?
3   Anders:  (3.0) Let's see (0.3) no wait a second (0.6) Now we   Stops, turns
             were actually within I think, and then (2.8) (Let's   Walks back
             see) (3.0) We were actually supposed to be a bit      Stops
             further down I think, but eh (0.7) we are at least
             allowed to pick up the next point here
4   Chris:   Hehe (0.8) Yess (0.6) Where's the river then?
5   Bård:    The most important thing is to win, not that we are
             supposed to learn so much?
6   Anders:  (Th) true. Eh. (0.8) It [depends, it depends if it
```

7	Bård:	[Maybe it is down there?	Points towards location
8	Anders:	>Maybe< (1.0) >You have picked up, eh, the point Bøkkerne, Message from Hesselberg: good work Premierløitnant Bielke, the barrels will be handy at sea< (0.5) Now you have most (0.4) of what you need to build the boats (0.4), .hh, as they are supposed to, drawings, money, ropes, hh, barrels, gunpowder, you only need to pick up cannons and cannonballs before you can go to the shipyard and get the construction started (0.8) Before you do that, it would be nice to quench your thirst, have a beer, [he he,	Reads from phone
9	Bård:	[Heh	
10	Anders:	.hh, and something to eat at Lars Evje's inn which is located (0.9) eh, along the sea in the direction of Åsane, yeah, I know where Åsane is.	Reads from phone Turns, walks

In turn 1 Anders is reading the GPS numbers from the phone as they walk. As we have noted, these numbers indicate their distance to the next location. By reading these numbers Anders makes the numbers available to the other participants and in this way makes them a potential resource in their joint orientation. Then he makes a joke by suggesting that the next location is a building in the vicinity that recently burned down — his last utterance in the form of laughter at the end of the turn further supports that this is meant as a joke. In turn 2, Bård asks when the incident took place, and by this indicates that he isn't aware of this — and thus accounts for his lacking response to the joke. The question remains unanswered.

In turn 3 Anders makes a change of topic and returns to the orientation. He says "Let's see, no, hang on". The first part of this utterance can be understood as an indication that he has some new information based on his reading of the display that contradicts their current movement in the given direction, (this is also made explicit by saying "no" in the middle of the sentence). In addition, this is a common way of saying that you are thinking or preparing the next statement. This is reinforced by the second part of the utterance "hang on", which also is a common way to make a bid for the next turn. While he is making this utterance he turns around and then begins to walk in the opposite direction right after he has finished talking. This point in turn 3 can also be seen as the start of the phase *arriving at the location.* Then Anders says, while walking, that he thinks they just were within the zone where they could pick up the next spot. This clarifies his abrupt change of direction. The other two participants turn around and follow Anders.

Now Anders has moved close to a fence and stops. He then says that they actually were supposed to be in another place. He does this by using the word "actually", and the chosen verb form "should have been" further indicates that it is someone's intention that they should have been at another location which is further down below the fence. Then the "but, eh" suggests that they will not do that. He then says that they can pick up the spot from where they are. Chris, in turn 4, asks where the river is. This refers to the text describing the next location they received at the previous location, where a river is mentioned. In this way the utterance is also in reference to the location they were supposed to be in, mentioned by Anders in turn 3. Simultaneously, Bård makes an ironic comment on their situation (turn 5). He says that the most important thing is to win, not to learn. This can be seen as a kind of evaluation of their predicament. They know they are supposed to be in another place, further down, below the fence, but they are allowed to pick up the spot in the game (since they are within the perimeter of the location, according to the GPS signals). This will save them time and thus increase the chances of winning the game. Anders answers "true" to the comment suggesting that he agrees and starts a statement that modifies this first agreement by saying "it depends". Bård simultaneously says that it might be down there, pointing down below the fence. Picking up the spot from where they are would save them the time and effort it would take to walk down to the physical location to which Bård is pointing. This would, due to the geography of the location, imply walking around several blocks. Anders repeats "it depends" but stops, and makes a short reply to Bårds statement, by saying "maybe". While they were discussing their location and where they were supposed to be, Anders has, obviously, picked up the spot on the mobile, and have now received the text. He then starts reading from the phone. This ends their discussion and removes any doubt as to whether they should actually go and find the physical location. In the last turn (10), he is reading from the phone that the next spot is in the direction of a suburb of Bergen, Åsane. He then turns around and starts walking, while saying that he knows where Åsane is.

Figure 2. Participants at Location.

As it is evident in this analysis, there are many things going on in this episode. We see how their joint orientation and movement involves both making the readings from the GPS available within the group, the use of bodily orientation and movement, and explicitly taking up and discussing where they are heading. Further we see an example of how they make some of the aspects of the game explicit and a topic of discussion (such as deciding on whether to find the physical location). We also see how the sequential structure of the talk and interaction is an important resource in the interactional organisation of the game play. It is also shown how this is a collaborative effort and that the game play relies on a number of contingent circumstances depending on how they navigate through the city streets and make use of the resources available to them, such as their knowledge of the city's geography and how they align this with the navigational resources offered through the phone's GPS and map, and the hints provided by the textual descriptions in the game.

Discussion

One key issue that is visible in the interaction and sheds further light on observations made in earlier phases of analysis in the project (see also Wake and Baggetun, 2009), is that there seems to be a contradiction between immersion in the game in the sense that winning the game is set up to use as little time and hints as possible, and time to dwell and reflect upon the historical aspects of the setting and the actual buildings and sites of historical significance. In the episode analysed above, the participants choose to move along without clearly identifying the physical location. Other groups, in other episodes in our material, spend more time and effort on finding the actual location, and are less oriented to move along quickly. This variation is, to a certain extent, dependent on the design of the game and the physical layout of the location in question, but ultimately something that is decided through the groups' interactional organisation of their game play. As is apparent in the analysis, Bård topicalizes this aspect in turn 5 in the excerpt. This statement is, as such, recognition of the dual purpose of the scenario. On the one hand, through this statement, the educational aim of the game is recognised; they are supposed to learn something about history. On the other they can approach this as a game where the important thing is to win. A similar finding is analysed by Hemming, Randall, Marr and Francis (2000), when they discuss how the "educational' character is 'there to be seen', and seen in the detail of activities" (Hemming, Randall, Marr & Francis, 2000, p. 224) of schoolchildren's visits to a museum. They look at how, even though there are clear educational goals and pedagogical inscriptions found in the arrangements and artefacts, the lessons to be learned "are contingent, in situ achievements of the parties to the interaction" (ibid., p. 234).

Another issue related to learning from the game is visible in turns 8-10 in the excerpt, where Anders reads out the text describing the mission to find the next location in the game. This text can be understood as instructions to the group, if understood and followed accordingly, will ensure the group's progress in the game. Suchman (2007), building on Lynch, Livingston and Garfinkel (1983), describes how the problem of following instructions "is to bring canonical descriptions of objects and actions to bear on the actual objects and embodied actions that the instructions describe" (Suchman, 2007, p. 111). In other words, turning the received, necessarily partial, instructions into something that can be recognised and practically accomplished – something that can be carried out. In the excerpt we see how Anders reads quickly through the narrative description and in and through this speedy reading checks that there is nothing they need to do with the information presented. Then in turn 10 when he recognises the referred place, Åsane, he immediately construes this, by stating that he knows where it is, as something that can be acted upon, and then turns around and walks, followed by the two other group members. In this way, the instructions are acted upon on and carried out through their embodied actions in the local environment.

Brown and Chalmers (2003) have discussed how way finding with various resources is a central part of tourism, and an activity through which tourists learn about the place they are visiting. This is relevant for the game play in this analysis. As much of the game play consists of activities, similar to those of the regular tourist, such as way finding and navigating in an urban landscape, an important aspect of the game is getting to know the city in a new manner – connecting the historical narrative to actual buildings and sites. This involves both engaging with the resources available through the game, but also bringing their knowledge of the local environment and geography to bear on the instructions and information presented in the game. Furthermore, the different characteristics of the urban, physical environment are used as game clues and information in order to complete the game. This was supported by the design of the game, by bringing hints about the physical environment into the game missions. In the analysed episode, however, the physical environment acted as a constraint in finding the location they were looking for, as it was visibly a long walk from where they were. Other episodes illustrate how participants used street-names, signs containing historical information set up by the local government of Bergen, and identifying characteristics of the environment referenced in the mission text as means of orientation in the game. This shows how the resources available in the game vary depending on the participants' navigation in the urban environment and how they in the practical accomplishment of the game play must deal with such a variation of contingent circumstances.

Concluding Remarks

This paper has presented a study of the interactional organisation and practical accomplishment of collaborative game play in a location-based game. Through analysis of the video-based material, the study has addressed how the participants in the gaming activity use the resources available to them in the game space, and make these resources available to each other as part of the practical accomplishment of game play. One of the practical aspects of accomplishing the game is related to navigation and way finding in the urban environment. A key issue that have been discussed is the apparent contradiction between the competitive features of the game and time to dwell on and reflect upon the historical aspects of the local environment as highlighted in the game, and how this was resolved at a practical level in and through the participants' interaction.

References

Baggetun, R., & Wake, J.D. (in preparation). Integrating location-based gaming with classroom activities.

Barab, S. & Squire, K. (2004). Design-based research: Putting a stake in the ground. *Journal of the Learning Sciences, 13*, 1-14.

Bennerstedt U., & Linderoth, J. (2009). The Spellbound Ones: Illuminating Everyday Collaborative Gaming Practices in a MMORPG. In: C. O'Malley, D. Suthers, P. Reimann, A. Dimitracopoulou (Eds.), *Computer supported collaborative learning practices: CSCL2009 Conference proceedings* (pp. 404-413). Rhodes, Greece, July 2009.

Brown, B., & Chalmers, M. (2003). Tourism and mobile technology. In: K. Kuutti, E. H. Karsten, et al. (Eds.), ECSCW 2003: *Proceedings of the eighth European conference on computer supported cooperative work*, Helsinki, Finland. Pp. 335-355, Dordrecht: Kluwer Academic Press.

Brown B., Chalmers, M., Bell, M., MacColl, I., Hall, M., MacColl, I., Rudman, P. (2005). Sharing the square: collaborative leisure in the city streets. *Proceedings of ECSCW 2005*, Paris, France. Pp. 427-429. Springer Verlag.

Brown, B., & Laurier, E., (2005). Maps & journeying: an ethnomethodological approach. *Cartographica*, *4*(3), 17-33.

Chang, B., Wang, H.-Y., Chen, C.-S., & Liang, J.-K. (2009). Distributed Weather Net: Wireless sensor network supported by inquiry-based learning. In: C. O'Malley, D. Suthers, P. Reimann, A. Dimitracopoulou (Eds.), *Computer supported collaborative learning practices: CSCL2009 Conference proceedings* (pp. 365-369), Rhodes, Greece. July 2009.

Derry, S.J., Pea, R.D., Barron, B., Engle, R.A., Erickson, F., Goldman, R., Hall, R., Koschmann, T., Lemke, J.L., Sherin, M.G., & Sherin, B.L. (2010). Conducting Video Research in the Learning Sciences: Guidance on Selection, Analysis, Technology and Ethics, *Journal of the Learning Sciences, 19*(1), 3-53.

Dourish, P. (2004). *Where the action is. The foundations of embodied interaction.* Massachusetts: MIT Press.

Goodwin, C. (1994). Professional vision. *American Anthropologist, 96*(3), 181-209.

Heath, C. & Luff, P. (2000). *Technology in action.* Cambridge: Cambridge University Press.

Heath, C., Hindmarsh, J., & Luff, P. (2010). *Video in qualitative research. Analysing Social Interaction in Everyday Life*. SAGE Publications Ltd.

Hemmings, T., Randall, D., Marr, L., Francis, D. (2000). 'Task, talk and closure: situated learning and the use of an 'interactive' museum artefact'. In S. Hester, D. Francis, (Eds.), *Local Educational Order*, (pp. 223–244). Amsterdam: Benjamins.

Ke, F. (2007). Using Computer-based Math Games as an Anchor for Cooperative Learning. Proceedings of the 8th iternational conference on Computer supported collaborative learning, pp. 354-356.

Koschmann, T., Stahl, G., & Zemel, A. (2007). The video analyst's manifesto (or the implications of Garfinkel's policies for the development of a program of video analytic research within the learning sciences). In R. Goldman, R. Pea, B. Barron & S. Derry (Eds.), *Video research in the learning sciences*. Retrieved from 28th Oct 2010.

Kirriemuir, J., & McFarlane, A. (2004). Literature Review in Games and Learning. Future Lab Literature Review, Report 8. Available at: http://www.futurelab.org.uk/resources/publications-reports-articles/literature-reviews/Literature-Review378

Klopfer, E., Perry, J., Squire, K., Jan, M.-F., & Steinkuehler, C. (2005). Mystery at the museum: a collaborative game for museum education. Proceedings of the 2005 conference on Computer support for collaborative learning, pp. 316-320.

Kurti, A., Milrad, M., & Spikol, D. (2007). Designing Innovative Learning Activities Using Ubiquitous Computing. *Proceedings of Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007),* (pp. 386-390).

Kurti, A., Spikol, D., Milrad, M., Svensson, M., & Petterson, O. (2007). Exploring how Pervasive Computing Can Support Situated Learning. In: *Proceedings of Pervasive Learning 2007*, (pp. 19-26).

Lynch, M., Livingston, E., & Garfinkel, H. (1983). Temporal order in laboratory work. In K. Knorr & M. Mulkay (Eds.), *Science observed* (pp. 205-238). London: Sage.

Lyons, L. (2009). Designing Opportunistic User Interfaces to Support a Collaborative Museum Exhibit, In: C. O'Malley, D. Suthers, P. Reimann, A. Dimitracopoulou (Eds.), *Computer supported collaborative learning practices: CSCL2009 Conference proceedings* (pp. 375-384). Rhodes, Greece, July 2009.

McFarlane, A., Sparrowhawk, A., & Heald, Y. (2002). Report on the Educational Use of Games: an Exploration by TEEM of the Contribution Which Games Can Make to the Education Process, Teem: Cambridge.

Roschelle, J., & Pea, R. (2002). A walk on the WILD side: How wireless handhelds may change computer-supported collaborative learning. *International Journal of Cognition and Technology*, *1*(1), 145-168.

Roschelle, J., Rosas R., & Nussbaum M. (2005). Towards a design framework for mobile computer-supported collaborative learning, *Proceedings of the 2005 conference on Computer support for collaborative learning*, Taiwan, July 2005, (pp. 520–524).

Rosenbaum, E., Klopfer, E., Broughner, B., & Rosencheck, L. (2007). Engaging Students in Science Controversy Through an Augmented Reality Role-Playing Game. Proceedings of the 8th iternational conference on Computer supported collaborative learning, pp. 608-612.

Satwitcz, T., & Stevens, R. (2007). Tools of Play: Coordinating Games, Characters, and Actions While Learning to Play Video Games. Proceedings of the 8th iternational conference on Computer supported collaborative learning, pp. 629-638.

Shute, V.J., Rieber, L., & Van Eck, R. (in press). . To appear in R. Reiser & J. Dempsey (Eds.), *Trends and issues in instructional design and technology, 3rd Edition*. Upper Saddle River, NJ: Pearson Education Inc.

Suchman, L. (2007). *Human-Machine Reconfigurations*. *Plans and Situated Actions*, (2 ed.). New York: Cambridge University Press.

Tan, T.-H., Liu, T.-Y., & Chang, C.-C. (2007). Development and Evaluation of an RFID-based Ubiquitous Learning Environment for Outdoor Learning. *Interactive Learning Environments, 15*(3), 253-269.

Wake, J.D., & Baggetun, R. (2009). "Premierløitnant Bielke": A Location-based Game for Teaching and Learning History. *International Journal of Mobile and Blended Learning, 1*(4), 12-28.

White, T. (2006). Code talk: Student discourse and participation with networked handhelds. *International Journal of Computer-Supported Collaborative Learning*, *1*(3), 359-382.

Yatani, K., Onuma, M., Sugimoto, M., & Kusunoki, F. (2004). Musex: A system for supporting children's collaborative learning in a museum with PDAs. *Systems and Computers in Japan*, *35*(14), 54-63.

Acknowledgments

SILO and the adhering mobile application builds on earlier development work done by Rune Baggetun on the MOTEL framework in the TRANSFORM project, and we would like to thank him for allowing us to build on this work. We thank Bjørge Næss for programming the SILO system and the mobile application. We would also like to thank Lars Wake for taking the pictures for Figure 1, and finally the participants that allowed us to film them while game playing.

Analysis of Small Group Interactions in a Seamless Language Learning Environment: An Artifact-Oriented Approach

Lung-Hsiang Wong, Wenli Chen, Ying Zhan, Learning Sciences Lab., National Institute of Education, 1, Nanyang Walk, Singapore 637616,
Email: lunghsiang.wong@nie.edu.sg, wenli.chen@nie.edu.sg, ying.zhan@nie.edu.sg
Chee-Kuen Chin, Singapore Centre for Chinese Language, 287, Ghim Moh Road, Singapore 279623, cheekuen.chin@sccl.sg

Abstract: We present a study in "Move, Idioms!", a mobile Chinese Language learning approach that emphasizes learner created content and contextualized meaning making with their daily encounters. Students used smart-phones on a 1:1, 24x7 basis to capture photos of the real-life contexts pertaining to Chinese idioms or conjunctions, made sentences with idioms/conjunctions, and posted them onto a wiki space for peer review. This paper focuses on students' on-campus face-to-face collaborative learning process. We derive a novel visualization approach for descriptive analysis of the small group activities inspired by the notions of mediation by artifacts and distributed cognition to provide a synoptic view of the process of student artifact co-creation in such collaborative activities. The artifact-oriented analysis and visualization approach is our preliminary attempt in making sense of how the seamless learning process may look like in the perspective of learners' individual and collaborative learning experiences.

Introduction

From e-learning to m-learning, the most publicly known phrase to describe these new advancements is perhaps "learning anytime, anywhere." The mobile technology enhances student learning whenever and wherever they are motivated to learn (Chan, et al., 2006). Whereas the rise of e-learning a decade ago had resulted in educators' concern of aggravating the digital natives' indulgence in the cyberspace, we argue that through proper m-learning design that emphasizes learners' interactions and meaning making with the peers (f2f) and the material reality (e.g., Rogers & Price, 2008; Wong, Boticki, Sun, & Looi, in press; Wong & Looi, 2010; Zhang, et al., 2010), the technology would instead bring them "back" to the physical world (Wong, 2010b).

Chan et al. (2006) define seamless learning as an approach where a student can learn whenever they are curious in a variety of scenarios and in which they can easily switch from one scenario or context to another (formal and informal learning, personal and social learning, physical world and cyberspace, etc.) using the personal device as a mediator. Nevertheless, after several years of relevant studies, the potential of mobile-assisted seamless learning has yet to be fully explored. One major challenge is that seamless learning has been blended into learners' day-to-day lives where every-"thing" within and beyond the four walls of the classroom has the potential of supporting or distracting/constraining their learning. This salient characteristic of seamless learning makes it difficult for any existing analytic tool to capture the full picture of seamless learning journey. Furthermore, the rich learning resources and contexts that could be distilled from learners' non-academic daily lives to complement formal learning are often neglected by both learners and teachers (e.g., Coffield, 2000).

How can learners' habit of mind and skills in making meaning with their daily encounters, and associating those with their formal learning gains be nurtured through their participations in (teacher-) *facilitated seamless learning (FSL)* (Wong, 2010a) experiences? This is one of the major research issues of our recent study in exploring a FSL design for Mobile-assisted Language learning (MALL), "Move, Idioms!". In the study, we facilitated a Primary 5 (11-year-old) class in Singapore to study 48 Chinese idioms (with 8 additional conjunctions in the later part of the study to experiment on the versatility of the learning design) over 10 months. Apart from in-campus idiom/conjunction lessons with contextualized and small-group learning activities (collaborative), the students were each assigned a smart-phone which they were allowed to access 24x7. With their smart-phones, they took photos in daily lives (personal), made sentences with the idioms/conjunctions, and then posted them onto a wiki space for peer review (collaborative). In analyzing the students' learning process in "Move, Idioms!", we are keen to uncover how the interplay of personal and collaborative meaning making (in both formal and informal settings, through non-ICT and ICT means), and the different forms of student artifacts constructed in different contexts, could create a "chaining" effect in mediating the students' learning journey (i.e., one piece of student work becomes a mediating artifact of the next stage of learning activity).

We have published (Wong, Chin, Tan, & Liu, 2010) our in-depth analysis on one particular form of student artifacts – the photo/sentence sets that they shared on the wiki, in the aspect of personal meaning making. In this paper, we shift our focus to the students' on-campus face-to-face (f2f) collaborative learning processes, not only with the aim of revealing how the social meaning making took place during their discussions with the mediation of various forms of student artifacts, but also establishing a construct for our future attempt to link the outcomes of such social learning experiences to their out-of-class, personal artifact creation and peer reviews.

Our rationale behind the design of the small-group f2f activities is to motivate and prepare the students for their out-of-school personal learning experiences. The small-group activities can be viewed as a group "exercise" of what and how the students can individually do in performing their out-of-class learning activities – closely observing and reflecting upon their living environments, associating environmental contexts (or creating contexts with the aid of physical objects or people) with their learned idioms/conjunctions, and generating artifacts for sharing. With this, we derive a novel visualization approach for descriptive analysis of the small-group activities of "Move, Idioms!' that is inspired by the notions of mediation by artifacts and distributed cognition. It is not (yet) our intention in this paper to formalize this approach for more general use, but more to provide a helicopter view of the process of student artifact co-creations in such collaborative FSL activities.

Mediation by Artifacts

Research findings show that classroom learning mediators include tasks, teacher and peer resources, subject content, and semiotic artifacts (e.g., languages, textbooks, PowerPoint and worksheets) (Liang, 2009). Artifacts (broadly defined to include instruments, signs, languages, and machines) mediate activity and are created by people to control their own behavior (Nardi, 1996). As (Stahl, 2002) posits, if we adopt a Vygotskian view of mediation by artifacts, then the knowledge construction process can be conceptualized as the construction of knowledge artifacts, involving physical and symbolic artifacts as starting point, as medium and as product.

The notion of mediation by artifacts, as formulated by Leont'ev (1981), accounts for material activity and its outcome in the form of transformed material objects. More recently, spoken and written discourse began to figure in the lists of mediating artifacts (Wells, 2002). Leont'ev (1974) argues, "A tool mediates activity that connects a person not only with the world of objects, but also with other people.." Distributed cognition offers a similar notion; e.g., Hutchins (1987) discusses "collaborative manipulation," the process by which we take advantage of artifacts designed by others (and ourselves), sharing ideas across time and space.

Combining both social and cognitive aspects, a distributed cognition perspective (Hollan, Hutchins, & Kirsh, 2002) suggests that cognitive activities such as knowledge construction are distributed across individuals and information artifacts through and with which they interact. The perspective implies the unit of analysis at the cognitive system as a whole, and emphasizes understanding of the coordination among the individuals and the artifacts in a system. Under the perspective, we would look for transformations of representations across individuals where those transformations can be interpreted as an intersubjective cognitive process. Examples include merging, revising, and connecting representations of ideas (Suthers, 2006).

We found the commonality between the small-group learning activities of "Move, Idioms!" and both notions of mediation by artifacts and distributed cognition. Our work is also related to an interpretative study (Alcock, 2005) on young children's co-construction of playful narrative events, motivated and mediated by artifacts – where Alcock put forward the concept of "distributed imagination" and claimed it an analogy of distributed cognition. One notable aspect in Alcock's thesis is that she treats people (e.g., the teacher who orchestrated the learning activity; a child who used her own body as an artifact to imitate a television character for her peers) as a potential form of artifacts. That echoes Cole's (1996) wider overarching concept of artifacts that people may be used as mediating objects. In addition, the meaning and use of artifacts are structured and transformed through activities. Hence, the term "mediating artifacts" is not necessarily in the traditional sense of (persistent) man-made objects, but could refer to any element (object or human) involved in the cognitive system of a learning activity, which can be appropriated into a mediating artifact that serves students' learning needs. Such a view is congruent with (Latour, 1996) argument that "to act is to mediate another's action" (p.237) – both humans and objects mediate, and one can only proceed to action by mediating another's action.

"Move, Idioms!" –Theoretical Framework and Learning Experience Design

As a fundamental component of language learning, vocabulary learning is often delivered in conventional ways, such as providing abstract definitions and sentences taken out of the context of normal use (Miller & Gildea, 1987). Such pedagogical strategies may pose a greater problem for learning of context-dependent vocabularies, such as conjunctions, idioms and proverbs. The complex nature of such vocabularies may result in highly context-dependent appropriateness of their usage. There are many possible real-life contexts where such vocabularies could suitably (or unsuitably but often mistakenly) be used, which are almost impossible to be prescribed in a simple definition.

Recognizing both the importance and the limitation of formal, in-class language learning, language learning theorists have been advocating the integrations of formal and informal (Titone, 1969), and personal and social (Noel, 2001) language learning, which mesh well with the notion of seamless learning. Informed by the theories, we developed a cyclic, customizable learning experience design of "Move, Idioms!" (see Figure 1).

The processes of the four activities are described below:

Activity 1 – In-class/campus contextual idiom learning: The activities are conducted to motivate and prepare students to engage in subsequent out-of-school activities. During each lesson, new idioms are introduced to the students in the means of multimedia presentations. The teacher then conducts contextualized learning activities such as facilitating the students to take photos in the campus to illustrate the idioms.

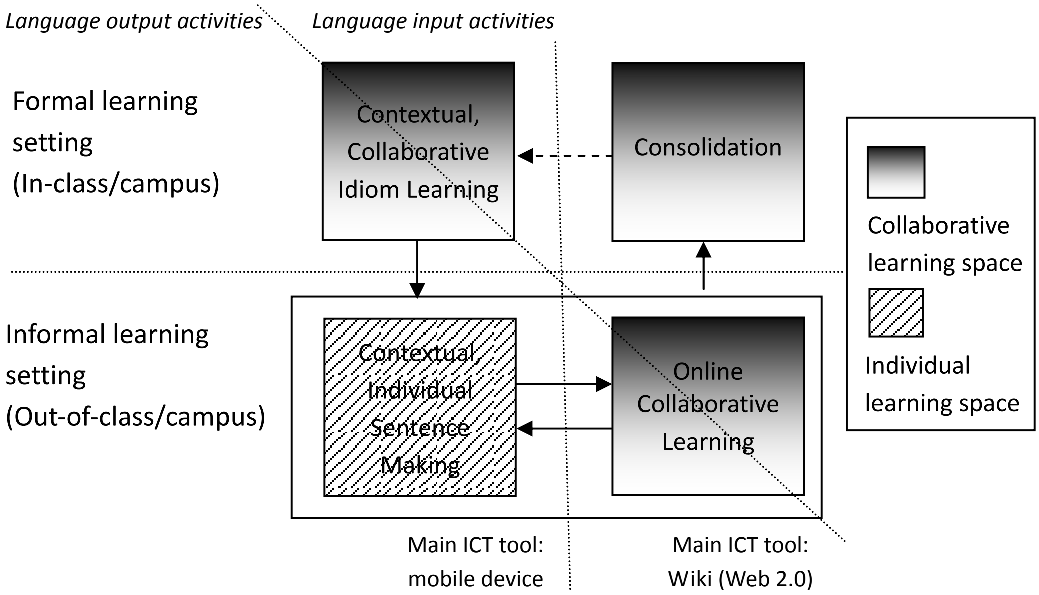


Figure 1. The "Move, Idioms!" Learning Experience Design.

Activity 2 – Out-of-class, contextual, independent sentence making: Students carry the mobile phones assigned to them 24x7 in order to identify or create contexts in their daily lives which could be associated with the idioms. They then take photos, make sentences by using the idioms to describe the photos, and post them onto a class wiki space. We create one wiki page for each idiom for students to post their artifacts. This offers convenience for comparing student-identified/created contexts pertaining to the same idioms.

Activity 3 – Online collaborative learning: Students perform peer reviews on the wiki by commenting on (with the built-in comment tool of wiki), correcting or improving their peers' sentences (by modifying the sentences posted on the wiki pages).

Activity 4 – In-class consolidation: Each student group is assigned a few existing student artifacts (photo/sentence sets) on the same wiki page with a mixture of correct, ambiguous and erroneous usages of an idiom. The groups compare the artifacts and revise the sentences if necessary. Subsequently, teacher-led classroom discussion makes contradictory views among the groups surfaced, facilitating class-wide debates.

The Enactment of "Move, Idioms!"

Our design research study of "Move, Idioms!" took place in January-October 2010. A class of 34 Primary 5 students, with mixed abilities in Chinese Language (as L2), participated in the study. Each of them was assigned a Samsung Omnia II smart-phone running MS Windows Mobile 6.5, with built-in camera, Wi-Fi access, Internet browser and English/Chinese text input. We (the researchers and a group of Chinese teachers) co-designed 8 "Activity 1" and 2 "Activity 4" lessons (see Table 1) which were then enacted by the Chinese teacher with 2-4 week intervals. With a graphic designer background, the teacher had had eight years of teaching experience and had been one of the ICT-mediated Chinese curriculum leaders in her school. Meanwhile, we featured off-the-shelf mobile-optimized comic animations that depict the meanings of the taught idioms during each lesson, which could also be accessed by the students anytime, anywhere.

An Artifact-oriented Approach to Analyze the Small-Group Artifact Co-creation Activities of "Move, Idioms!"

In our attempt to unpack the learning processes of the small-group activities during the in-class lessons of "Move, Idioms!", one particular aspect that comes to our attention is the role of artifact mediation throughout their rich, situated learning experience, which can be characterised as an experience situated in 'continually moving and re-constructed contexts' (Looi, Wong, & Song, in-press). In turn, we carried out descriptive analysis on the verbatim transcriptions of audio and video recordings as well as the field notes of student group interactions during all the in-class lessons. The 8 student groups were comprised of 4-5 members each, which were randomly assigned by the teacher prior to the study and were heterogeneous in Chinese Language competencies. We adopt the above-stated perspective of mediating artifacts as the analytical means. The intention of the analysis is to identify various forms of learning support tools and intermediate products, both physical and non-physical, map them into the artifact-oriented perspective, and unveil their relationships as well as various paths that different student groups had taken to accomplish their learning tasks in different lessons. We engaged student groups to member-check our analysis – though we did not use the academic term "artifacts" and instead asked them to confirm and further elaborate "things that contributed to or distracted their photo/sentence co-creation activities" that we discovered. Due to the space constraint, we will not go into the detailed analysis but to present the synoptic view of our findings.

Table 1: Summary of 10 "Move, Idioms!" in-class lessons (all were "Activity 1" lessons unless otherwise stated).

Lesson ID: #1	Worked in groups of 4, students brainstormed and made sentences that utilized the idioms (one idiom per sentence) learned in this lesson and sketched the scenario on paper worksheet. Students were not assigned the smart-phones yet.
#2	Students were assigned the smart-phones after receiving technical training. Worked in groups, they repeated the activity of Lesson #1 except that they were asked to enact the scenarios and took photos within the classroom. They then signed out the phones for 24x7 access.
#3	Each student group was assigned a particular area within the campus (e.g., canteen, basketball court, ecological garden). They went to the designated area, brainstormed to associate their encounters with their learned idioms, took photos and made sentences. They were encouraged to make one sentence that utilized two idioms.
#4	Each group brainstormed a paragraph that utilized multiple (both newly learned and previous) idioms. In-class Photo taking was not allowed but they were asked to plan for four pictures to depict the group-generated scenario.
#5	Each group brainstormed a paragraph that utilized multiple idioms, and took photos within the campus.
#6	An "Activity 4" lesson.
#7, #8, #9	Same as #5. For Lesson #8 and #9, they were taught and brainstormed sentences that utilized learned conjunctions instead – although most sentences had incorporated some idioms as well.
#10	An "Activity 4" lesson.

The unit of analysis is the group photo/text co-creation session. By artifact we refer not only to the student products (photo/text sets) but also the mediating artifacts and emergent, intermediate products involved in their learning processes. We classified the artifacts identified in the analysis into four categories. The classification is based on the major role of each artifact in the learning process. To simplify the analysis, we exclude task regulating artifacts such as timers and teacher's monitoring but focus on identifying artifacts that are directly contributing to the contents of the students' outcome artifacts. The four types of artifacts are:

- Physical artifacts: Physical or environmental tools that mediate the learning activities, such as the ICT tools, the classroom and the campus (and the physical objects available there), and even people (teachers, students, researchers, guests) who posed for photo shooting;
- Subject matter artifacts: Artifacts that represent the target knowledge to learn – the idioms and conjunctions themselves (which are linguistic/cultural artifacts), and the example sentences, paragraphs, or photo/text sets given by the teachers (digital artifacts – in PowerPoint form), as well as comic animations and YouTube videos, to demonstrate the usage of the vocabularies;
- Socio-cognitive artifacts: Non-physical artifacts generated through socio-cognitive means, such as teacher's verbal scaffolds and peer discourses (both are semiotic artifacts), students' *in-situ* improvising or emergent strategies to carry out the learning activities, ideas and stories for photo taking and sentence/paragraph compositions, and their shared (evolving) understandings in the associated meanings and linguistic functionalities of the idioms and conjunctions;
- Outcome artifacts: The target artifacts that the students are explicitly required to generate, including photos and text, peer review comments and peer revised text.

Hatch and Gardner (1993) propose a concentric model of the forces affecting a distributed cognition system. The three forces in the model are personal, local and cultural. Our level of analysis with the four types of artifacts being identified fits to the local level. We generated diagrams which depict all the artifact-mediated processes leading to outcome artifact creations as we have observed in the small group f2f discussions. Due to the space constraint, we present only the diagram pertaining to the combination of Lesson #5, #7, #8, #9 (with a photo set and a coherent plot in a paragraph as the outcome artifact) in this paper (see Figure 2). Although it is possible to generate one diagram per group per lesson, we decided to feature one consolidated diagram per lesson design for the same reason (space). The arrows in the diagrams represent mediatioon-outcome relationships. Each numbered node in the diagrams denotes a mini-state, which we refer to as *joint mediation*, of multiple artifacts (linked by incoming arrows) to the student group that results in certain output artifacts (linked by outgoing arrows). We refer to such nodes as joint mediation nodes (JM-nodes). Note that for any JM-node with multiple "input" artifacts, it is not necessarily that all (but can be any combination of) the "input" artifacts will be utilized in every instance of the joint mediation. The same goes for a JM-node with multiple output artifacts (not necessarily all but can be any combination of output artifacts). The numbers in the JM-nodes show the rough sequence of different joint mediations (not definitive, as some groups could have occasionally backtracked to previous states to revise their earlier artifacts) in the co-creation processes. The JM-nodes labeled with the same number plus an alphabet (e.g., 1a) represent supplementary joint mediations that occurred in the same mini-state and were likely to be interweaving (or never happened in some cases). In addition, we identified certain mediating artifacts that occasionally did not play their usual constructive roles and instead became distracting or constraining factors to students' co-creation processes. We refer to them as *constraining mediation*.

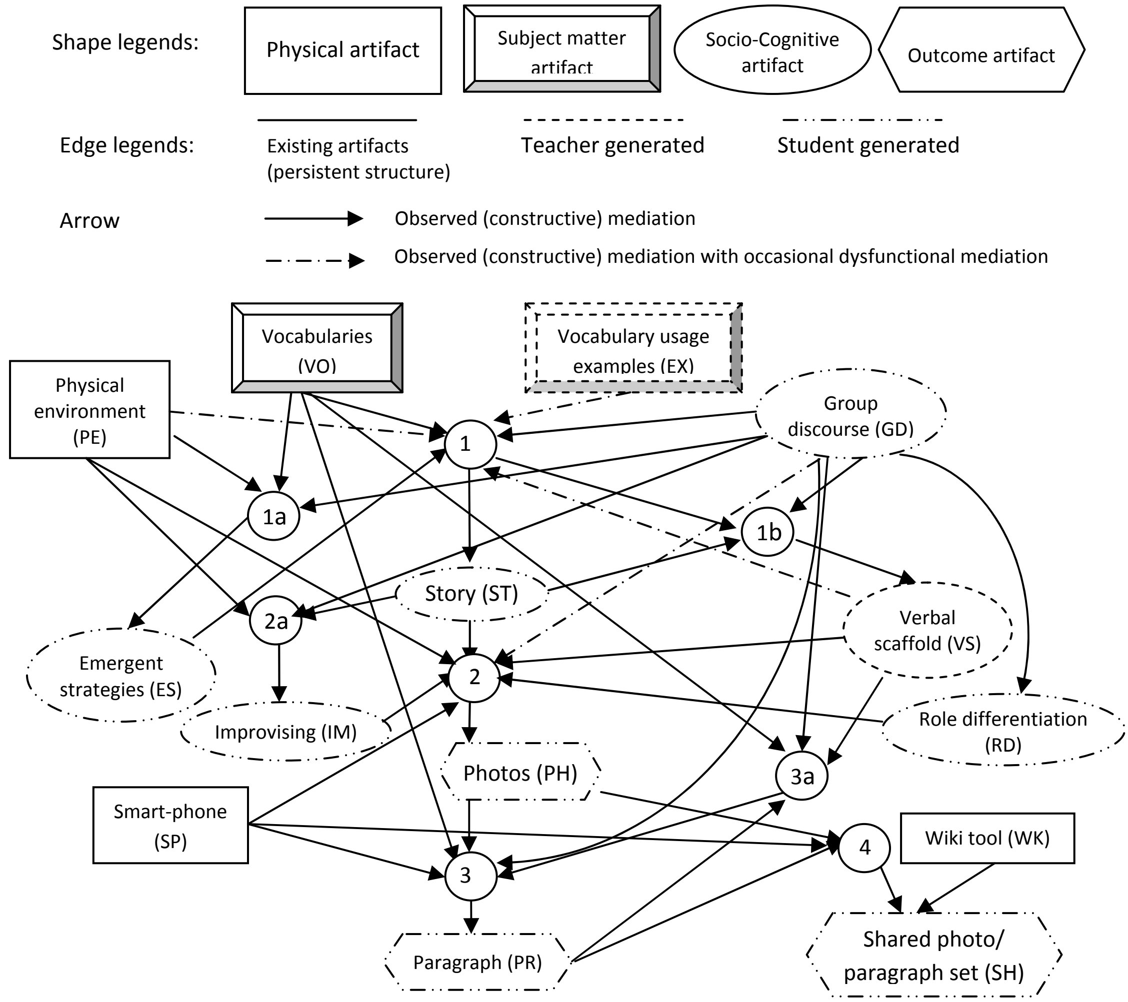


Figure 2. The Artifact-oriented Diagram of Collaborative Learning Activities in Lesson #5, #7, #8, and #9.

Analysis of Small-Group Artifact Co-creation Processes in Lesson #5, #7, #8 and #9

The small-group activities in Lesson #5, #7, #8 and #9 can be divided into four mini-states – story co-generation (the result of the joint mediation of 1), photo set co-generation (the result of 2), paragraph co-generation (the result of 3), and wiki page appending (the result of 4).

JM-node 1 & 1a: Fresh from learning new idioms/conjunctions, student groups brainstormed their stories in the class, with the set of vocabularies to use (VO), the vocabulary usage examples (EX), the potentially accessible physical environment (PE; see the next paragraph for elaboration), and the (emergent) group discourse (GD) and teacher's verbal scaffolds (VS) as mediating artifacts. In particular, the vocabulary usage examples (EX) were presented by the teacher prior to the small group activities, in the forms of idiomatic/conjunction animations, PowerPoint presentations of photo and sentence/paragraph sets, or videos. Such digital artifacts had inspired or influenced some groups' subsequent storytelling (ST). In many cases, we consider the EX a constraining mediation as the artifacts constrained the students' creativity – some student groups copied the essential story idea and only made minor changes in the characters or the props. That is, they were not able to apply the learned idioms in different contexts, thus limiting their deep learning and internalization of the idioms. However, there were also positive cases where student groups managed to work out stories inspired by the EX but with more enriching contents and utilizations of different sets of idioms.

Even with the "comfort" of brainstorming (GD) stories (ST) in the classroom rather than *in-situ*, students usually took into consideration the accessibility of the physical environment (PE) for their subsequent photo shooting. This might instead become a constraining mediation as they had to revise the story and even drop some of the idioms (VO) in mind to satisfy the PE constraints. For example, a student proposed using the idiom 鸟语花香, literally means "birds twitter and fragrance of flowers" but figuratively characterizes "a fine, beautiful day", in their story. The group spent minutes to debate about how tedious it would be to take a photo with both twittering bird(s) and flowers in it within the campus. Even though a member aptly pointed out the metaphorical nature of the idiom and could be used to describe fine weather and pleasant sceneries, with or without actual birds and flowers in the context, his group-mates did not concur and eventually gave up the idiom.

Conversely, for another group, the PE constraints triggered them to figure out strategies (ES) to overcome those. For example, another group which was similarly stuck in how to portray鸟语花香 (a fine, beautiful day) had the teacher granted permission to take photos after school. They then worked out a plan of taking four photos and assigned each group member the task to take one of them, with a student who claimed that she could take a photo with birds and flowers from her home balcony assuming the obvious responsibility.

The common misconception on the usage of鸟语花香 (a fine, beautiful day) was later clarified by the teacher to the class at another lesson. However, we argue that through such socio-cognitive learning activities, students were able to more effectively internalize the correct functionality of the idiom after the clarification. Therefore, such a mediating artifact which is seemingly constraining the students' outcome artifact creation in short term may turn out to be conducive in their learning in a long run if the teacher is able to scaffold for "making errors work for the students and not against them" (Rubin & Thompson, 1982).

<u>JM-node 1b</u>: During the small group activities, the teacher was usually "touring" among the groups to check their progress and provide verbal scaffolds (VS) to improve their contextualized stories (ST) where necessary. Therefore, the VS is usually a product of the joint mediation of VO, PE and/or GD. However, there were occasional cases where the teacher advised against certain student groups' story ideas for no pedagogical reason. One possibility was that the teacher's designer background had made her subconsciously more 'interventionist' in students' creative processes and products. We consider that another form of constraining mediation that might have distracted the students' creativity.

<u>JM-node 2 & 2a</u>: The joint mediation of ST, PE, GD and the smart-phone (SP) resulted in the production of the first part of an outcome artifact – the photo set (PH). Improvising (IM) prompted by PE, including the sets (e.g., the library) and props (e.g., putting math and science books, sketch paper, a ruler and a calculator on the desk of an inventor character in the story – these props were contributed by different students) available, took place in some photo shooting occasions that resulted in the adaptation of their original ST. All the group-generated ST required some group members to become actors and enacted the scenario. Therefore, most of the groups were getting self-organized (i.e., role differentiation or RD) with dedicated "directors" and "photographers" being appointed (or self-appointed), and they sometimes switched roles between directors, photographers and actors. There were also cases where co-directing and co-photographing took place and due to the simplicity of the storyline, they still managed to carry out the learning task smoothly.

The smart-phone (SP) did not *only* serve as a productive tool. Occasionally, students checked a photo on the phone display immediately after shooting, and decided whether a retake was needed to make sure their idea was correctly executed and the idiom association was appropriate. In turn, SP became a cognitive tool (albeit still a physical artifact) to mediate their deeper thinking.

<u>JM-node 3</u>: The joint mediation of PH, VO, GP and SP resulted in the production of the second part of an outcome artifact – the paragraph (PR). Unlike during JM-node 1 where student groups only worked out rough story ideas (ST) that "guaranteed" usage of some of the given idioms or conjunctions, this would be the time that they synthesized their ST and photos taken (PH) and threaded the idioms they had in mind together to become a coherent paragraph (PR). In developing their paragraphs, some groups re-looked at the photos taken by different students (although most groups would have appointed an "official" photographer during JM-node 2, some students had taken extra photos when they were not engaged in acting) and replaced or inserted photos to the original photo sets. Additional idioms might incidentally be added to the paragraph. One instance took place in Lesson #5 where a group came back to the classroom from photo shooting on a story about basketball playing. A student recalled that her group-mate who played the role of basketball player (and became a PE artifact) was sweating minutes after the shooting started, which was captured by the photos (PH). She asked if 汗流浃背 ("all of a sweat") is an idiom (VO) and a group-mate confirmed that. They decided to incorporate the idiom into the paragraph (PR). We checked with the group afterward and found out that the idiom was not taught in the present year but two of the group members recalled that their Chinese teacher in the previous year taught it in the class. The other two members who came from another class with a different teacher had not come across the idiom before; but they now learned it from their group-mates.

<u>JM-node 3a</u>: During JM-node 3, some groups found it difficult to organize their photos, or organize the idioms and their stories into paragraphs. The teacher was again "touring" among the groups and provided the

much needed scaffolding (VS). Typically, she looked at the unfinished paragraphs (PR) and offered ways to organize the relevant artifacts, sometimes with additional or alternative idioms (VO) being proposed.

JM-node 4: These were merely synthesizing-and-posting tasks (synthesizing photo and sentence into a complete instance of outcome artifact, and posting it onto the web) although they did append new contents to the relevant wiki pages, i.e., a new version of a wiki page, which was a collective outcome artifact, was generated.

Discussion

Seamless learning is probably one of the most complex forms of learning as it involves multi-facets of learners' daily lives and has the potential of integrating most of the ICT-enhanced learning models – digital classroom learning, e-learning, m-learning, context-aware ubiquitous learning, etc. A holistic seamless learning experience design requires that learners not only interact with other people and instructor-provided artifacts within a relatively closed learning environment (e.g., traditional classroom or e-learning portal), but also with the authentic physical environment and perhaps the Internet at large, where learners may draw any element that they incidentally encounter and appropriate it into a useful mediating artifact for learning. We need additional analytic methods to unpack the individual and collaborative learning experiences of seamless learners. The most common existing CSCL analytical tools may offer a microscopic view of the dynamics of student collaborations within the situated learning context, most likely in a socio-cognitive constructivist perspective, but not so much on how individuals or groups interact with the environments. The artifact-oriented analysis and visualization approach reported in this paper is our preliminary attempt in making better sense of how the seamless learning process may look like, in the perspective of seamless learners' individual and collaborative learning experiences.

We are not the first who adopted such an artifact-oriented approach in analyzing learning environments, i.e., to map relevant elements in such environments into mediating artifacts for subsequent analysis. Apart from the above-stated analysis on young learners' narrative co-construction activities (Alcock, 2005) which bears the greatest resemblance with our work, there had been other studies (e.g., Conole, 2008; Lei, 2008; Wang & Ching, 2003) which adopted similar approaches of artifact-oriented analysis.

The uniqueness of our work is twofold. First, we developed a visualization approach to reveal the interdependence among the mediating and outcome artifacts within relatively complex learning processes perhaps with multiple branching – as the flexibility of multiple learning pathways (see: Looi, et al., 2009) is a significant feature for seamless or authentic learning environments. In our diagrams, we observed what Stahl (2002) conceptualizes as "… the construction of knowledge artifacts, involving physical and symbolic artifacts as starting point, as medium and as product." (p.67) Artifacts changed their roles through learners' appropriation for carrying out different learning tasks. Seamless learners ought to assume greater agency in deciding what and how to learn, either personally or collaboratively and across different learning spaces, rather than always being "dictated" by the externally-supplied mediating artifacts with prescribed roles. In turn, their habit of mind and skills of identifying and appropriating artifacts (including their previously learned knowledge and previously created artifacts) to mediate both their planned and incidental learning would become crucial for them.

Second, the above-stated literature almost only portrayed ideal situations where all mediating artifacts worked well in supporting student learning. As artifacts, in nature, both enable and constrain human activities (Paavola & Hakkarainen, 2004), we identified some weak links (constraining mediations) of the learning processes through our analysis on the empirical data – artifacts which may instead constrain or distract the learning tasks under certain circumstances. We believe that such findings (together with the identified positive emergent strategies and teacher's scaffolds) could be used to inform future pedagogical and even technological (re)design to eliminate or reduce such constrains and strengthen the use of emergent positive artifacts.

Conclusion and Future Work

This paper reports on our effort in analyzing the small-group f2f artifact co-creation processes in the "Move, Idioms!" learning environment. We derived an artifact-oriented visualization approach, inspired by the notions of "mediation by artifacts" and distributed cognition, to meet this end. Due to the strong nature of situated learning in our design, analyzing the discourses among the group members leading to their outcome artifact co-creations without taking into consideration their interactions with the environment is inadequate. We believe such an approach can also be applied to analyze the students' personal, out-of-school learning experiences, especially that the small-group artifact co-creation activities were designed in the way as a preparation of their personal artifact creation activities in informal settings. Subsequently, the interplay between collaborative and personal learning can be investigated. Still, more work need to be done to formalize the approach, such as a more rigorous categorization of artifacts, the similarities and differences in applying the approach to analyze collaborative and personal learning experiences, and how can it be applied to inform future learning design.

References

Alcock, S. J. (2005). *A socio-cultural interpretation of young children's playful and humorous communication.* PhD Dissertation, Massey University, Palmerston North, New Zealand.

Chan, T.-W., Roschelle, J., Hsi, S., Kinshuk, Sharples, M., Brown, T., et al. (2006). One-to-one technology-enhanced learning: An opportunity for global research collaboration. *Research and Practice in Technology-Enhanced Learning, 1*(1), 3-29.

Coffield, F. (2000). The structure below the surface: reassessing the significance of informal learning. In F. Coffield (Ed.), *The Necessity of Informal Learning* (pp. 1-11). Bristol: Policy Press.

Cole, M. (1996). *Cultural Psychology: A Once and Future Discipline*. Cambridge, MA: Harvard University Press.

Conole, G. (2008). Capturing practice: The role of mediating artefacts in learning design. In L. Lockyer, S. Bennett, S. Agostinho & B. Harper (Eds.), *Handbook of Research on Learning Design and Learning Objects: Issues, Applications and Technologies* (pp. 187-207). Hershey, PA: IGI Global.

Hatch, T., & Gardner, H. (1993). Finding cognition in the classroom: an expanded view of human intelligenc. In G. Salomon (Ed.), *Distributed Cognitions* (pp. 164-187). New York: Cambridge University Press.

Hollan, J., Hutchins, E., & Kirsh, D. (2002). Distributed cognition: Toward a new foundation for Human-Computer Interaction research. In J. M. Carroll (Ed.), *Human-Computer Interaction in the New Millennium* (pp. 75-94). New York: ACM Press Addison.

Hutchins, E. (1987). Metaphors for interface design *ICS Report 8703*. La Jolla: University of California, San Diego.

Latour, B. (1996). On interobjectivity. *Mind, Culture and Activity, 3*(4), 228-245.

Lei, X. (2008). Exploring a sociocultural approach to writing strategy research: Mediated actions in writing activities. *Second Language Writing, 17*, 217-236.

Leont'ev, A. N. (1974). The problem of activity in psychology. *Soviet Psychology, 13*(2), 4-33.

Leont'ev, A. N. (1981). The problem of activity in psychology. In J. V. Wertsch (Ed.), *The Concept of Activity in Soviet Psychology* (pp. 37-71). Armonk, NY: Sharpe.

Liang, X. (2009). Investigating mediations in student activities in an English immersion context in mainland China. *English Language Teaching, 2*(4), 38-50.

Looi, C.-K., Wong, L.-H., So, H.-J., Seow, P., Toh, Y., Chen, W., et al. (2009). Anatomy of a mobilized lesson: Learning my way. *Computers & Education, 53*(4), 1120-1132.

Looi, C.-K., Wong, L.-H., & Song, Y. (in-press). Discovering Mobile Computer Supported Collaborative Learning. In C. Hmelo-Silver, A. O'Donnell, C. Chan & C. Chinn (Eds.), *The International Handbook of Collaborative Learning*. New York: Routledge.

Miller, G. A., & Gildea, P. M. (1987). How children learn words. *Scientific American, 257*, 94-99.

Nardi, B. A. (1996). Studying context: A comparison of activity theory, situated action models, and distributed cognition. In B. A. Nardi (Ed.), *Context and Consciousness: Activity Theory and Human-Computer Interaction* (pp. 35-52). Cambridge: MIT Press.

Noel, K. A. (2001). New orientations in language learning motivation: Towards a model of intrinsic, extrinsic, and integrative orientations and motivation. In Z. Dörnyei & R. Schmidt (Eds.), *Motivation and Second Language Acquisition* (pp. 43-68). Honolulu: University of Hawaii, Second Language Teaching and Curriculum Center.

Paavola, S., & Hakkarainen, K. (2004). *"Trialogical" processes of mediation through conceptual artifacts*. Paper presented at the Scandinavian Summer Cruise at the Baltic Sea 2004, Stockholm, Sweden.

Rogers, Y., & Price, S. (2008). The role of mobile devices in facilitating collaborative inquiry in situ. *Resarch and Practice in Technology Enhanced Learning, 3*(3), 209-229.

Rubin, J., & Thompson, I. (1982). *How to be a More Successful Language Learners*. Boston: Heinle and Heinle.

Stahl, G. (2002). Contributions to a theoretical framework for CSCL. *Proceedings of the Conference on Computer Support for Collaborative Learning* (pp. 62-71), Boulder, USA.

Suthers, D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. *Computer Supported Collaborative Learning, 1*(3), 315-337.

Titone, R. (1969). Guidelines for teaching second language in its own environment. *The Modern Language, 53*(5), 306-309.

Wang, X. C., & Ching, C. C. (2003). Social construction of computer experience in a first-grade classroom: Social processes and mediating artifacts. *Early Education and Development, 14*(3), 335-361.

Wells, G. (2002). Dialogue in activity theory. *Mind, Culture and Activity, 9*(1), 43-66.

Wong, L.-H. (2010a). From facilitated seamless learning to self-directed seamless learning. *Proceedings of the Global Chinese Conference on Computers in Education 2010* (pp. 33-40), Singapore.

Wong, L.-H. (2010b). *Theme-based Invited Speech (CUMTEL)*. Paper presented at the International Conference on Computers in Education 2010, Putrajaya, Malaysia.

Wong, L.-H., Boticki, I., Sun, J., & Looi, C.-K. (in press). Improving the scaffolds of a mobile-assisted Chinese character forming game via a design-based research cycle. *Computers in Human Behavior*.

Wong, L.-H., Chin, C.-K., Tan, C.-L., & Liu, M. (2010). Students' personal and social meaning making in a Chinese idiom mobile learning environment. *Educational Technology & Society, 13*(4), 15-26.

Wong, L.-H., & Looi, C.-K. (2010). Vocabulary learning by mobile-assisted authentic content creation and social meaning-making: Two case studies. *Computer Assisted Learning, 26*(5), 421-433.

Zhang, B., Looi, C.-K., Seow, P., Chia, G., Wong, L.-H., Chen, W., et al. (2010). Deconstructing and reconstructing: Transforming primary science learning via a mobilized curriculum. *Computers & Education, 55*(4), 1504-1523.

Digital Media in the Classroom: A Study on How to Improve Guidance for Successful Collaboration and Learning in Student Teams

Carmen Zahn, Karsten Krauskopf, Friedrich W. Hesse, Knowledge Media Research Center, Konrad-Adenauer-Strasse 40, 72072 Tuebingen, Germany,
Email: c.zahn@iwm-kmrc.de, k.krauskopf@iwm-kmrc.de, f.hesse@iwm-kmrc.de
Roy Pea, Stanford Center for Innovations in Learning (SCIL), Stanford, CA, USA, roypea@stanford.edu

Abstract: Digital video technologies provide a variety of functions to support collaborative knowledge construction. Yet, for novice learners, positive outcomes also depend on effective guidance of group interactions. In this paper, we present empirical evidence for the use of web-based video tools to support students' collaborative learning in a history class. In an experiment with 16-year old learners (N=148) working with a history topic, we compared two contrasting types of guidance for student collaboration in dyads (cognitive task-related guidance or social interaction-related guidance). We also compared two types of video tools. Both types of guidance and tools were aimed at supporting students' active, meaningful learning and critical reflection. Results indicate that social interaction-related guidance was more effective in terms of learning outcome (e.g., critical reflection skills) than cognitive task-related guidance. The different tools did not yield significant differences in learning. The practical implications of these results are discussed.

Introduction

Computational technology and digital media can greatly enhance the possibilities for creative knowledge construction in social learning situations. However, there are open questions related to guidance of group interactions in desirable directions, especially when novice learners face complex authentic learning tasks: For example, a major concern expressed from the instructional perspective is how instructive guidance should be designed in accordance with human cognitive functioning (e.g., Kirschner & Sweller, 2006). In addition, CSCL research has emphasized the necessity of considering the complex relations between tasks, tools, interaction processes and learning outcomes (e.g., Van Drie, Van Boxtel, Erkens, & Kanselaar, 2003). In our contribution, we tap into these issues by examining the example of digital video technologies used for collaborative knowledge construction in a classroom setting. Specifically, we investigate in an experiment *how* instructive guidance can be balanced for middle-school students in order to support skill-intensive socio-cognitive processes during a short collaborative design task for history learning with different digital video tools.

The potential of digital video technologies reaches far beyond the dynamic presentation and illustration of visual information. With digital video tools, learners may zoom into and out of digital video sequences, they may insert hyperlinks into videos in order to relate visual information to other instructional materials, and they may arrange video sequences for discussion and reflection. Such functions are expected to afford, for example, detailed observations (e.g., Smith & Reiser, 2005), multiple perspectives (e.g., Goldman, 2004) or the understanding of complex concepts in ill-structured domains (Spiro, Collins, & Ramchandran, 2007). The affordances of digital video technologies can be restructured for youthful learners in classrooms, so that students can either create their own representations (e.g., multimedia documents) or arrange video contents in order to understand and explain complex subject matter (Zahn, Pea, et al., 2005). This usage, in the sense of *learning through design* (e.g., Kafai & Resnick, 1996), goes far beyond teacher-centered approaches where videos on curriculum topics are only watched by individual learners or in whole-class models.

Over the last several years, we have investigated collaborative design with video tools. Evidence from our experimental studies has indicated that specific affordances of video tools (e.g., of WebDIVER™, Pea, et al. 2004), when employed in design tasks for history learning, can support learners' social interactions to become more productive than those performed with simple technological solutions, resulting in improved learning outcomes (e.g., Zahn, Pea, Hesse & Rosen, 2010). Yet, initial field studies with 16-year-old students (Zahn, Krauskopf, Hesse, & Pea, 2010) showed that the positive effects of video tools were sometimes limited to an "action-level", and students would have needed more guidance to optimize their collaborative design process. This finding is consistent with findings from Barron (2003) showing that student groups can have problems engaging in productive knowledge-building conversations during video-based mathematics problem solving. It is also consistent with related evidence showing that collaborating students need help in organizing, planning and conducting scientific inquiries (Edelson, Gordin, & Pea, 1999), in scientific argumentation (Kollar & Fischer, 2004) and in accomplishing scientific design projects (Kolodner, et al. 2003).

Two sources of problems can hinder productive socio-cognitive processes when students perform design tasks with digital video tools: The complexity of collaboration with *video tools* and the complexity of collaborative *design.* We have demonstrated in prior research how specific video tools can influence collaborative learning (e.g., Zahn, Pea, Hesse & Rosen, 2010). In the present study, we take into account their *differential* complexity (Zahn, Pea et al. 2005) when they are used as design tools for learning. *Design* tasks generally consist of creating and structuring content for an anticipated audience according to the aesthetic standards of the media at hand. They include the setting of design goals and complex processes of knowledge transformation, as was proposed earlier by related cognitive research (e.g., Bereiter & Scardamalia, 1987, Goel & Pirolli, 1992; Hayes, 1996). According to Détienne (2006) *collaborative* design includes the management of task interdependencies and of multiple perspectives. Correspondingly, design activities relate to the levels of the design problem/design solution and group cooperation. Moreover, when designers use complex and sometimes unfamiliar *digital tools* (video tools in our case), they coordinate their collaboration by establishing a social problem space that is distributed over the cognitive systems of at least two people *and* a digital artifact, creating new coordination problems familiar in distributed cognitive systems (Streek et al., 2011). Based on this shared context, they negotiate their choices of design goals and their understanding of content, task schemas, genre knowledge, and task relevant strategies (as in *collaborative writing*, e.g., Lowry, Curtis, and Lowry, 2004). The importance of the shared (multimodal) context for design was repeatedly emphasized (Détienne, 2006).

Consequently, although designing video or other artifacts with digital tools is highly desirable for students, because it is cognitively engaging, students may sometimes be cognitively overwhelmed by the complexity of having to find a design solution, manage the group and use an unfamiliar digital tool. They actually may need guidance throughout the process so that learning through design can take place. Based on previous research on the nature of design (e.g., Détienne, 2006), we might provide such guidance, tackling either *cognitive design task*-related issues or *social interaction*-related issues (similar to Fischer et al.'s (2002) distinction of *content-specific* and *content-unspecific* instructional support or Weinberger et al.'s (2005) *epistemic* vs. *social scripts*). It is still open whether guiding students' design activities or their social interactions would lend important support for successful task completion - or whether students might feel restricted by too much guidance and be impeded in their creativity and learning. Also, the mediating role of the digital video tools for collaboration under such conditions is quite unclear. Hence, in our study, we compared the two forms of guidance using two types of video tools, and we explored whether interactions would occur.

Experimental Study

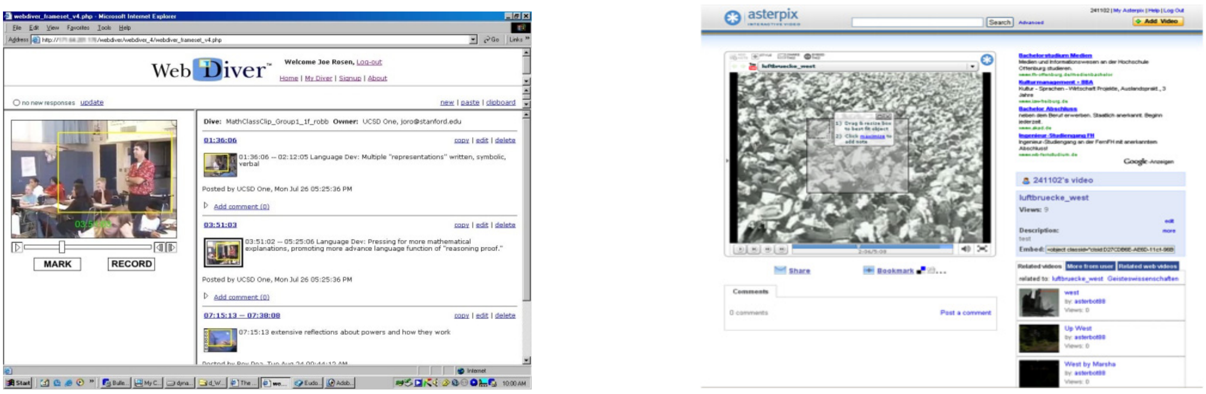
Method

Participants: 148 students (81 male, 65 female, 2 no answer) from four different German high schools located in Southwestern Germany participated in the study. Their mean age was M = 16.2 years (SD = 1.0). Prior to the study we obtained written consent from the students' parents and the school administration. The sample size varies minimally (see Tables 1 to 3), due to problems with data availability from stored design products and video taped interactions.

Study design: The study was conducted in a computer classroom set up at our institute. Classes accompanied by their respective teachers came to the institute on regular school days and as part of their regular history curriculum. Upon arrival they were randomly grouped into dyads and assigned to one of the four experimental conditions of a 2 x 2 study plan. The first factor *Guidance* (cognitive design-related vs. social interaction-related) determined which type of instructive guidance was provided to support the collaborative accomplishment of a visual design task: guidance either emphasizing the cognitive aspects of the design task (e.g., setting a design goal, planning a design concept, tailoring information *for an audience*), or guidance focusing on smooth collaboration (e.g., developing *common* ground about design goals, and design decisions, determining *communication* rules for discourse practice). The second factor *Video Tool* determined whether the students worked with WebDIVER™ (Pea, et al., 2004) or Asterpix as their design tool: The tools differed on a generic level in either supporting collaborative analysis (WebDIVER tool for guided noticing) or collaborative linking of information (Hypervideo tool Asterpix). With WebDIVER, learners' cognitive/collaborative analysis is heightened by their ability to zoom into and out of digital video sequences, and arrange digital video sequences for discussion and reflection. With the Hypervideo tool Asterpix, the collaborative ability to insert new knowledge artifacts into an existing digital video is heightened by hyperlinks relating visual information to other materials. All other circumstances were kept constant across conditions.

Task: A visual design task based on a historical newsreel was employed. This task had been carefully developed for the purpose of studying computer supported history learning with digital video tools in a realistic classroom (e.g., Zahn, Krauskopf, Hesse & Pea, 2010). It follows central educational goals in the domain of history in German middle school education (Krammer, 2006). Furthermore, it is theoretically founded in cognitive and collaborative frameworks of advanced learning and knowledge building approaches (e.g., Scardamalia, 2002). During this task, students work on a newsreel about the Berlin blockade in 1948, so that it can be published, e.g., on a website of a virtual history museum. They are asked to analyze and comment on the newsreel so that future visitors of the virtual museum can develop a good understanding of both the content and the style of the newsreel as a propaganda instrument. To accomplish the task, the students can use a collaborative video tool (see Tools section). The constructive activity of designing content for a web page of a virtual history museum provides learners with a framework for comparison and re-organization of knowledge, as they produce their own ideas and work creatively with them. During the collaborative design process, it is assumed that learners appropriate the video content to their own thinking purposes and develop advanced thinking skills. The learning goal and a special challenge for the students is to understand that the newsreel is not only "showing" the history topic (Berlin 1948), but that the newsreel itself *is* a history topic (i.e., a newsreel as an historical means for propaganda). In other words, historical content knowledge is closely intertwined with developing advanced thinking skills (Scardamalia, 2002), like being able to analyze and critically reflect on video messages.

Materials and Tools: The video used in the visual design task is a digitized version of an historical newsreel originally produced by the Allied forces (US/Great Britain) and shown to the German public during the Berlin blockade in 1948. It covers news information about the airlift established in 1948 by the Allied forces when Russia tried to cut off Berlin from traffic of goods. It consists of 95 single pictures and lasts five minutes. The video used in the transfer task is a modern 65-second TV-Clip by the German Green Party (Buendnis 90/Die Gruenen) from the 2006 nationwide election in Germany. The texts used in the experiment contain 350-1500 words each. The content of the texts provides detailed information on three sub-topics: accounts of the historical context of Berlin in post-war Germany, information on media history and newsreels in post-World War II Germany, and a short introduction on film theory. Guidance was implemented in text-based form within the computer environment used for general task instruction. The texts differed between conditions in their descriptions of how one should best proceed to solve a visual design problem. The video tool used for computer-supported learning in the visual design task was either WebDIVER (see Figure 1a) or Asterpix (see Figure 1b). WebDIVER is one of the software programs developed in the DIVER Project (http://diver.stanford.edu) at Stanford University. It is based on the metaphor of enabling a user to "dive" into videos for expressing points of view regarding precise spatio-temporal video areas of one or more source videos. Asterpix is a commercially available hypervideo tool. It is based on the idea of enabling users to select areas of interest and place graphical hyperlinks into a source video.



Figures 1 a and b. Graphical user interfaces of the collaborative (hyper-)video tools used in the study: (a) WebDIVER^{TM} (right), (b) Asterpix (left).

With the functions offered by WebDIVER, users can select either a temporal segment or a spatio-temporal sub-region of a video by mouse-controlling a rectangular selection frame (acting like a camera viewfinder) to "pan" and/or "zoom" into view only that subpart of a video that they wish to feature, and then interpretively annotating their selection via a web interface. Each dive movie clip and its associated annotations is represented in a panel in the dive, and a remix of the video clips and annotations can be played to experience the dive. Asterpix was a Web 2.0 tool (http://www.asterpix.com/, no longer available) with functions based on the hypervideo idea: Users could isolate dynamic, sensitive regions within video materials, provide text commentaries to these regions and add links to other web resources. The links could further be discussed by means of an integrated e-communication tool. Thus, users could include their own annotations and knowledge in a video and share them with others in a group or community (cf. Zahn et al., 2005).

Procedure: A week before the students came to our lab, they filled in questionnaires that assessed their prior knowledge and other control variables. The *experimental* sessions consisted of the following steps: In Step 1 (preparation phase), the students individually read the overall instructions, including the different types of guidance (either guidance for effective design or guidance for effective social interactions during design). Then they read the history/media texts, and watched the video showing the historical Berlin-Blockade newsreel from 1948. They briefly practiced the use of the video tools to establish familiarity. In Steps 2, 3 and 4 (collaborative design and learning phase) the participants worked collaboratively in dyads at a computer. In Step 2 (planning), those students in dyads in the *social interaction*-related guidance condition were asked to write down the content they would like to cover in their design products and how they would like to coordinate their design work. Those students in dyads in the *cognitive design*-related guidance condition were asked which design goals they would aim for. In Step 3, the dyads were asked to design their product according to their initial ideas using either WebDIVER or Asterpix. In Step 4 (evaluation) the dyads were asked to evaluate the quality of their own products and teamwork. When students were done, they continued with Steps 5 and 6 (test phase), where self-assessment questionnaires and knowledge tests were completed individually. In Step 7, the participants individually accomplished a transfer task (TV-ad, see Materials section). They were then thanked and released and went back to their schools with their teachers. During the whole procedure, the teachers were present and tutors were also available for any questions or technology problems.

Measures: To assess prior background knowledge in the domain of history, computer expertise or expertise in film and media production, a pre-questionnaire (self-assessment) and a multiple choice knowledge test were administered. To assess the effectiveness of our text-based instruction as implementation of guidance (manipulation check), we asked the subjects to select a maximum of three alternatives from six statements about the task's characteristics (3 social characteristics, e.g., "one of the most important aspects of the learning unit was good communication" and 3 design characteristics, e.g., "one of the most important aspects of the learning unit was to design for a target audience"). To assess collaborative design performance, the design products created by the dyads with WebDIVER or Asterpix were analyzed. From these products, the following categories of data were obtained: "video selections/sensitive areas with comments", "style features commented", and "interpretations" in the comments. Additionally, dyadic interactions were captured with a webcam and a screen recorder (Camtasia Studio by TechSmith). The proportions of talking time in the categories "design planning", "design action", "design evaluation", "technical issues", "problems", and "off task" (related to total amount of talking time) were extracted from these video data using video analysis software (Videograph©). To assess treatment effects on learning outcomes, a post-test was administered, consisting of a multiple choice test measuring historical topic knowledge and a transfer task tapping advanced visual analytic skills. The multiple choice post-test consisted of 8 items. (Sample item: "At the beginning of 1946 Germany is... a) ...a unified nation, b) ...divided into four sectors, c)... divided into an Eastern and a Western part, d) ...divided into 16 Länder." The theoretical maximum of this test was 13 points, and it had a satisfactory internal consistency, Cronbach's $\alpha = .71$. The transfer task part of the post-test was assigned to reveal skills of critical analysis and reflection in response to a video message. It consisted of two questions relating to a political TV-ad from the 2006 nationwide German government elections. ("Please analyse the following video sequence by answering the questions 1) Which film techniques were used? 2) What might have been the intention of using them?"). The questions were open ended.

Results

We will first present results substantiating the comparability of our conditions, and then results obtained from quantitative analyses of the design products and post-tests. Due to assumed interdependence of students working in one dyad, we determined dyads as the unit of analysis and used data aggregated within dyads (cf. Kenny, Kashy & Cook, 2006). The level of significance for all analyses was set to .05.

Comparability of the conditions: 2 x 2 between subjects ANOVAs with the factors Guidance and Video Tool revealed no significant differences between the conditions concerning participants' age, prior experience with computers in general and video software in particular, their history grade, or their dispositional interest in history (all $p > .10$). The dyads also did not differ significantly between conditions concerning within-group composition related to age, gender, prior knowledge, history grade, or historical interest (all $p > .10$). In addition, student dyads did not differ in their appraisal of the task, the appraisal of their teamwork or the amount of invested mental effort during task work (all $p > .10$), indicating that the participants' overall positive attitudes towards task and performance were similarly high in the four conditions. In sum, the conditions can be considered comparable. However, historical knowledge showed a marginally significant interaction, $F(1, 68) = 3.86$, $p = .05$, partial $\eta^2 = .05$, showing that for students working with WebDIVER, those participating in the cognitive design-related guidance condition scored higher on the pretest ($M = 10.23$, $SD = 2.55$) than students in the social interaction-related condition ($M = 8.22$, $SD = 2.20$), $t(34) = 2.53$, $p = .02$. For students working with Asterpix, there were no significant differences. All ANOVAs reported here were also run as ANCOVAs controlling for interest in history and prior knowledge, and they are reported when they show different results.

Manipulation Check: The means and standard deviations of students' choices in the question tapping their understanding of the task are shown in Table 1. An ANOVA revealed no significant difference between conditions concerning their scores in "design task" characteristics, Fs < 1, *ns*, but a significant difference for the "social task" characteristics for the factor Guidance, $F(1, 68) = 15.51$, $p < .001$, partial $\eta^2 = .19$. More "social task" items were chosen by students who had received social interaction-related guidance than by students who had received cognitive task-related design guidance. Our text-based implementation of guidance by task instructions can thus be considered effective for eliciting the students' awareness of the design problem in all conditions and the students' increased awareness of the *social* demands of the collaborative design task in the social interaction-related conditions.

Table 1: Means (*M*) and Standard Deviations (*SD*) for the manipulation check.

	Selective video editing tool (WebDIVER™)				Integrative video editing tool (Asterpix)			
	CDG (n = 18)		SIG (n = 18)		CDG (n = 19)		SIG (n = 17)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Manipulation check – social	0.81	0.75	1.37	0.65	0.71	0.49	1.31	0.58
Manipulation check – design	1.44	0.65	1.30	0.49	1.40	0.63	1.40	0.62

Note. CDG = cognitive design related guidance, SIG = social interaction related guidance

Design Performance: The means and standard deviations of the scores of the dyads' design products concerning numbers of commented video selections, style features and interpretations are presented in Table 2. Interrater reliability for style features and interpretations were satisfactory, Cohen's $\kappa \geq .94$. ANOVAs revealed a significant main effect for the factor Guidance: The mean scores in all the mentioned indicators were significantly higher for the products of dyads in the condition with social interaction-related guidance, than for those from dyads in the condition with cognitive design-related guidance, in terms of *number of comments*, $F(1, 67) = 6.46$, $p = .01$, partial $\eta^2 = .09$, *number of style features*, $F(1, 67) = 4.78$, $p = .03$, partial $\eta^2 = .07$, and *number of interpretations*, $F(1, 67) = 4.63$, $p = .04$, partial $\eta^2 = .07$. Hence, design performance in the visual design task was higher in the social interaction-related guidance conditions than in the other conditions. No further main or interaction effects were found. Thus, the two forms of video tools were not used in different ways - at least in relation to the quantitative indicators of design performance we applied here.

Table 2: Means (*M*) and Standard Deviations (*SD*) for the quality indicators of students' design performance.

	Selective video editing tool (WebDIVER™)				Integrative video editing tool (Asterpix)			
	CDG (n = 18)		SIG (n = 18)		CDG (n = 19)		SIG (n = 14)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Number of commented video selections	4.11	3.38	6.61	3.03	4.11	2.38	5.43	3.65
Number of style features	0.14	0.48	1.22	2.26	0.29	0.77	0.64	1.17
Number of interpretations	0.11	0.47	0.89	1.53	0.32	0.82	0.64	1.15

Note. CDG = cognitive design-related guidance, SIG = social interaction-related guidance

Historical Topic Knowledge: Analyses of the scores from the multiple choice post-test on knowledge about the history topic revealed a total mean score $M = 7.54$ ($SD = 2.46$) out of 13 possible points. We conducted a mixed 2 x 2 x 2 ANCOVA with the two between-subjects factors Guidance and Video Tool and the within-subjects factor Pre-Post-Test to test for differences in the gain in individual factual knowledge. After controlling for the differences in pre-test scores, the results still showed a significant increase in factual knowledge over time, $F(1, 67) = 34.80$, $p < .001$, partial $\eta^2 = .34$. However, there were no significant differences between the conditions, Fs < 1, *ns*, and no significant interaction, $F(1, 67) = 1.93$, $p = .17$ indicating that the students in both conditions had developed an understanding of the historical content.

Critical Analysis and Reflection: The students' written answers to the transfer task questions were coded independently by 2 raters. For the coding procedure, coders considered a pre-defined default solution created by an expert in visual media production (first author of this paper). The solution comprised exemplary stylistic features used in the TV-ad (e.g., camera, music, montage), as well as examples for correct interpretations of such elements (e.g., close-up of a person's face aims at creating emotional involvement). Based on this example, raters counted the number of named style features and interpretations. Also, the elaborateness of the answers was rated on a 3-point Likert scale (1 = simple, 3 = elaborate). Interrater reliability was satisfactory for the number of style features, Cohen's $\kappa = .91$, and the elaborateness rating, Cronbach's $\alpha = .80$. However, rater agreement for the number of interpretations of these style features was very low, Cohen's $\kappa = .10$. Closer analyses revealed that the raters differed greatly with regard to how strictly they applied the coding scheme. For further analyses we decided to only use the coding of the more conservative rater. The analysis of the transfer test results revealed a total average of $M = 1.97$ ($SD = 0.74$) for "number of style features", $M = 0.37$ ($SD = 0.23$) for "number of interpretations" and $M = 1.19$ ($SD = 0.47$), and for "elaborateness of the answer". ANOVAs revealed that the means (see Table 3) of all these indicators were significantly higher in the answers of students in the conditions with social interaction-related guidance, than in the conditions with cognitive design-related guidance: *number of style features*, $F(1, 68) = 7.96$, $p = .01$, partial $\eta^2 = .11$, *number of interpretations*, $F(1, 68) = 4.36$, $p = .04$, partial $\eta^2 = 06$, *elaborateness of the answer*, $F(1, 68) = 4.11$, $p = .047$, partial $\eta^2 = .06$. Overall, effect sizes were of medium to large size. There were no effects of the video tool factor, Fs < 1.1, *ns*, or any significant interactions, Fs < 1, *ns*. Thus, although all students developed a comparable understanding of the topic, the learning outcomes in terms of advanced thinking skills (critical analysis and reflection) were higher when social interaction was supported in the student dyads.

Table 3: Means (*M*) and Standard Deviations (*SD*) for the tests tapping factual knowledge and indicators of the transfer task.

	Selective video editing tool (WebDIVER™)				Integrative video editing tool (Asterpix)			
	CDG[b] (*n* = 18)		SIG[c] (*n* = 18)		CDG (*n* = 19)		SIG (*n* = 17)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Transfer test – critical analysis and reflection								
Number of style features	1.77	0.63	2.37	0.51	1.72	0.78	2.06	0.87
Number of interpretations	0.34	0.23	0.43	0.26	0.30	0.21	.42	0.19
Elaborateness of the answer	1.09	0.35	1.31	0.40	1.08	0.44	1.31	0.63

Notes. [a]Theoretical maximum = 13. [b]CDG = cognitive design-related guidance, [c]SIG = social interaction-related guidance.

Dyadic Interactions: For analyses of dyadic interactions, we coded the proportions of time students were engaged in activities belonging to one of the categories "design planning", "design action", "design evaluation", "technological issues", "problems" and "off task" (related to total amount of talking time, $M = 21.52$ minutes, $SD = 4.46$). 20% of the videos were coded by a second rater and agreement was on average satisfactory, median of Cohen's $\kappa = .64$. However, 2 x 2 ANOVAs with the two between-factors Guidance and Video Tool yielded no significant differences between the conditions (see Figure 2). Further analyses on a more fine-grained level are ongoing and will be available at the time of the conference.

Discussion

Our results provide evidence from an experimental study that helps to answer the question of how to improve guidance for student teams solving a complex authentic design task for history learning with the support of web-based video tools. Results indicate that while using either of the advanced video tools we offered was generally effective, differences in the types of guidance we implemented (*cognitive task*-related vs. *social interaction*-related guidance) resulted in different learning outcomes. Firstly, the immediate design products of the dyads' task work were of better quality. Secondly, individual students scored significantly higher in a transfer test evaluating critical analysis and reflection skills. Concerning factual knowledge about the topic ("Berlin blockade"), no differences and no trade-off effects in performance in a multiple-choice posttest emerged. Moreover, during the students' dyadic interactions, similar amounts of time were devoted to the subtasks "design planning", "design action", "design evaluation", "technical issues", "problems" and "off task" behavior in all conditions. Thus, the differences in the transfer test were neither at costs of other learning outcome measures, nor could they be explained by a first (superficial) analysis of specific students' interaction time spent on task. This finding was not confined to a specific tool used in our study: Results show that given the conceptual differences of the video technologies (WebDIVER and Asterpix) described above, the benefits of supporting the social problem space persist. We thus conclude that the dyads with social interaction-related guidance learned *more* than the dyads with cognitive task-related guidance, and we conjecture that even given different affordances for the two video tools, social interaction-related guidance improved the quality of dyadic interactions on a deeper content level. And this leads us to the question of *how exactly* that quality was improved. In a next cycle of analyses we will investigate differences in the content of dyadic interactions and be able to present the first results by the time of the CSCL conference. These findings will add further answers to the question of *how* instructive guidance can be balanced for middle-school students in order to support skill-intensive socio-cognitive processes.

When interpreting the results reported here to draw conclusions for school practice, we need to consider the following issues: In this study we created a highly controlled, computer-supported experimental setting, thereby enabling us to draw causal conclusions. We exposed students to a short-time visual design task for a regular history lesson, which is different from large design projects performed over several weeks. So the results cannot be generalized to such projects. However, we compared our results from this experiment with the results from an earlier field study in a real, "noisy" classroom situation with a comparable sample of students and with the same short task and test items (Zahn, Krauskopf, et al. 2010). Results revealed general gains in factual knowledge (pre- to post-tests) similar to those obtained in the field. No indications of influences of the artificial experimental situation (positive or negative) were found. From the analyses, we may thus conclude that students of the age group investigated here seem to have sufficient working patterns for completing short design tasks (establishing a design problem space), but not necessarily for social interaction (establishing a social problem space). This might be the case because design tasks are often used in school-based education and students are familiar enough with them to perform the necessary activities. However, they seem to be less able to activate effective ways of team interaction from their everyday school experiences. In other words: Guidance repeatedly emphasizing the aspects of design problem solving thereby focusing on the design product may *not* improve the learning addressed here, but guidance improving collaborative activity (coordinating teamwork and communication) can. For design-based interventions such as this, the result may be somewhat unsurprising, but certainly worth highlighting. The strength of the social interaction-related guidance described here is such that it calls for further analysis across a broad range of collaborative learning environments. For teachers this issue would be important in practice if, indeed, their guidance of students' collaborative task work in real lessons were focused on social interaction processes. This perspective is consonant with related views across different domains and digital media (e.g., Barron, 2003) – and hopefully stimulates further CSCL research.

References

Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences 12* (3). 307-359.

Bereiter, C., & Scardamalia, M. (1987). The Psychology of Written Composition. Hillsdale, NJ: Lawrence Erlbaum Associates.

Détienne, F. (2006). Collaborative design: Managing task interdependencies and multiple perspectives. *Interacting with Computers*, *18*(1), pp 1-20.

Edelson, D.C., Gordin, D.N. & Pea, R.D. (1999). Addressing the Challenges of Inquiry-Based Learning Through Technology and Curriculum Design. *The Journal of the Learning Sciences, 8* (3&4). 391-450.

Fischer, F., Bruhn, J., Gräsel, C. & Mandl, H. (2002). Fostering Collaborative Knowledge Construction with Visualization Tools. *Learning and Instruction 12.* 213-232.

Goel, V., & Pirolli, P. (1992). The structure of design problem spaces. *Cognitive Science, 16*, 395-429.

Goldman, R. (2004). Video perspectivity meets wild and crazy teens: Design ethnography. *Cambridge Journal of Education, 2*(4), 147–169.

Hayes, J.R. (1996). A new model of cognition and affect in writing. In M. Levy & S. Ransdell (Eds.), The Science of Writing. (pp. 1-27). Hillsdale, NJ: Erlbaum.

Kirschner & Sweller (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist Discovery, Problem-Based, Experiential, and Inquiry-based Teaching. *Educational Psychologist, 4*(2). 75-86.

Kafai, Y. B., & Resnick, M. (Eds.). (1996). *Constructionism in Practice: Designing, Thinking, and Learning in a Digital World* (pp. XII, 339 S.). Mahwah, NJ: Lawrence Erlbaum Associates.

Kolodner, J. L., Camp, P. J., Crismond, D., Fasse, B., Gray, J., Holbrook, J., et al. (2003). Problem-Based Learning Meets Case-Based Reasoning in the Middle-School Science Classroom: Putting Learning by Design™ Into Practice. *Journal of the Learning Sciences*, 12, 495-547.

Kollar, I., Fischer, F., & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. Learning & Instruction, 17(6), 708-721.
Lowry, P. B., Curtis, A., & Lowry, M. R. (2004). Building a Taxonomy and Nomenclature of Collaborative Writing to Improve Interdisciplinary Research and Practice. Journal of Business Communication, 41, 66-99.
Pea, R., Mills, M., Rosen, J., Dauber, K., Effelsberg, W., & Hoffert. E. (2004, Jan-March). The DIVER™ Project: Interactive Digital Video Repurposing. *IEEE Multimedia, 11*(1), 54-61.
Pea, R. D. (2006). Video-as-data and digital video manipulation techniques for transforming learning sciences research, education and other cultural practices. In J. Weiss, J. Nolan, J. Hunsinger, & P. Trifonas (Eds.), *International handbook of virtual learning environments* (pp. 1321–1393). Dordrecht, The Netherlands: Kluwer Academic.
Scardamalia, M. (2002). Collective Cognitive Responsibility for the Advancement of Knowledge. In B. Smith (Ed.), Liberal Education in a Knowledge Society (pp. 67-98). Chicago: Open Court.
Smith, B., & Reiser, B. J. (2005). Explaining behavior through observational investigation and theory articulation. *Journal of the Learning Sciences, 14*, 315–360.
Spiro, R. J., Collins, B. P., & Ramchandran, A. (2007). Reflections on a post-Gutenberg epistemology for video use in ill-structured domains: Fostering complex learning and cognitive flexibility. In R. Goldman, R. D. Pea, B. Barron, & S. Derry (Eds.), *Video research in the learning sciences* (pp. 93–100). Mahwah, NJ: Erlbaum.
Streeck, J., Goodwin, C., & LeBaron, C. (2011, in press) (Eds.). *Embodied interaction: Language and body in the material world*. Cambridge: Cambridge University Press.
Van Drie, J., Van Boxtel, C., Erkens, G., and Kanselaar, G. (2003). Supporting Historical Reasoning in CSCL. In B. Wasson, R. Baggetun, U. Hoppe, & S. Ludvigsen (Eds.), *Designing for Change in Networked Learning Environments. Proceedings of the International Conference on Computer Support for Collaborative Learning - CSCL 200*3, (pp. 93-102). Bergen, NO: InterMedia.
Weinberger, A., Ertl, B., Fischer, F., & Mandl, H. (2005). Epistemic and social scripts in computer-supported collaborative learning. *Instructional Science, 33*(1), 1-30.
Zahn, C., Pea, R., Hesse, F. W., Mills, M., Finke, M., & Rosen, J. (2005). Advanced Digital Video Technologies to Support Collaborative Learning in School Education and Beyond. In T. Koschmann, D. Suthers, & T.-W. Chan (Eds.), *Computer Supported Collaborative Learning 2005: The Next 10 Years* (pp. 737-742). Mahwah, NJ: Lawrence Erlbaum.
Zahn, C., Pea, R., Hesse, F.W. & Rosen, J. (2010). Comparing Simple and Advanced Video Tools as Supports for Complex Collaborative Design Processes. *The Journal of the Learning Sciences. 19*(3). 403-440.
Zahn, C., Krauskopf, K., Hesse, F.W. & Pea, R. (2010). Digital Video Tools in the Classroom: How to Support Meaningful Collaboration and Critical Thinking of Students? In: M.S. Khine & Saleh, I.M. New Science of Learning: Computers and Collaboration in Education. New York: Springer.

Weather Gods and Fruit Kids – Embodying Abstract Concepts Using Tactile Feedback and Whole Body Interaction

Carolina Johansson, Zeynep Ahmet, Jakob Tholander, Franc Aleo, Mobile Life Centre, Forum 100,
164 40 Kista, Sweden,
Email: lina@mobilelifecentre.org, zeynepa@mobilelifecentre.org, jakob@mobilelifecentre.org,
franc@mobilelifecentre.org
Martin Jonsson, Saiful Sumon, Södertörn University, 141 89 Huddinge, Sweden,
Email: martin.jonsson@sh.se, saiful.islam.sumon@sh.se

Abstract: In this paper we present findings based on the design and study of a game like activity that allows for physical and bodily interaction around abstract concepts like energy and energy consumption in a collaborative learning setting. The game, called Weather Gods and Fruit Kids, uses motion sensing technologies in combination with tactile and audio feedback to create an embodied interactive setting without computer screens. We analyze and discuss the properties of the interactive setting as well as the interactions with and around the system using characteristics such as multiple modalities of response, large space interaction and aspects of focus and attention. The work suggests that alternative pedagogical activities can be created providing new entries to theoretical concepts using an embodied interaction approach. In particular it may support kinesthetic learners in their preference to learn by being physically engaged.

Introduction

The use of physical and tangible objects has been a longstanding approach in design of technology for learning. A common argument both in learning technology design and human-computer interaction has been that physical interaction supports making abstract notions more concrete and easy to grasp (Ullmer & Ishii, 2000). A more recent trend is to design for interaction and experiences that involve bodily action and movement in a more significant way than by only introducing tangible interaction elements. This trend is sometime referred to as whole body interaction. Letting people engage in interactions using their whole bodies provides way to let people have new forms of experiences with technology that could not be achieved in a traditional hand-eye interaction styles.

In this paper we explore a game like activity called the Weather Gods and Fruit Kids designed for bodily interaction and kinesthetic learning in a school setting. The technology is designed to support children to interactively experience consumption, preservation and creation of energy, by sensing and responding to different aspects of the children's movements. Here we will present the design of the game, the design explorations leading up to it and some key outcomes from a study of children using the game.

Bodily interaction and the involvement of body and movement to interact with technology is an important part in the shift towards so-called post-WIMP interfaces. It concerns how by moving away from the traditional mouse and keyboard interaction we can allow for a richer, more multifaceted way of interacting with technology. By opening up the interaction space to include our full body we make use of a bigger range of our perceptual resources to make sense and understand the world (Klemmer et al., 2006). A range of recent work in HCI has addressed how bodily interactions and experiences should be understood and designed for, examples include focus on the felt aspects of experiences (Larssen et al., 2007) and approach of designing for the aesthetic dimension of physical experiences (Schiphorst, 2009). To understand such processes of experience some researchers have taken inspiration from physical activities with non-digital devices such as golf and skateboarding, or even horseback riding in order to outline key dimensions of physical experiences (see Tholander & Johansson, 2010; Höök, 2010).

The focus of the findings presented here is on the different ways children use their bodies in the interaction, the kind of experiences they have, and the role the technology plays in these processes. In the analysis presented we use the three themes from the analysis by Jonsson et al. on interactive spaces for physical interaction; (i) the role of the embodied performance for others; (ii) the role of the physical space in the interaction, and (iii) the role of device specific qualities in the unfolding of the interaction (Jonsson et al., 2009).

Background

There are the numerous explorations of technology that illustrate new ways of increasing bodily involvement in interaction, such as the Lega device for leaving physical traces of experiences at an art hall (Laaksolahti et al., 2010) , the eMoto system for bodily emotional expression (Sundström et al. 2007), BodyBeats for dance-based music creation (Zigelbaum et al., 2006), BodyBug the interactive movement partner (Moen, 2005) and the emotion and movement controlled game Emroll (Zangouei et al., 2010). Bodily ways of interacting have also grown commercially and become particularly popular in computer games. This includes an array of dancing and sports games. For instance in *Dance Dance Revolution* where the player follow a sequence of instructions by dancing on a sensor mat, in WiiSports sports where the player performs movements mimicking those required in real games of bowling, boxing, tennis, or in music games like *Guitar Hero* and *Rock Band* where the player acts out pieces of music by playing on replicas of real instruments. These technical developments have drawn attention to the need to further understand people's interaction and experiences with bodily engaging technology, as well as to the interplay between physical, social and digital aspects of our experiences.

Issues of physicality, materiality, and social space are central in research on embodied interaction, pervasive computing and tangible interfaces, in which the particular physical manifestation of a computational artifact and its consequences for people's interaction and learning with and through the artifact is commonly brought to discussion (Hornecker & Buur 2006; Klemmer et al., 2006). These issues are also becoming increasingly relevant for design of collaborative technology for learning (Price, Sheridan et al., 2010) through the emphasis on how cognitive processes are deeply intertwined with and dependent upon our physical bodies and our relations to the material and physical properties of our social world. We experience and make sense of the world by moving in it, by exploring new situations through bodily interaction and with the material, physical and social surroundings (Lakoff & Johnson, 1999; Säljö, 2000). Thus, new interactive technologies that support mobile, bodily, and social activities beyond the desktop setting enable a shift both in the development of new forms of user interactions and experiences, as well as new tools to be used in learning activities. This aligns with much CSCL research as not being only the study of how people interact and learn with designed artifacts, but also the study and design of the situations in which efficient learning is taking place (Price & Rogers, 2004).

This paper presents findings from our close collaboration with an elementary school in Stockholm, where we made use of bodily interaction in a kinesthetic learning context. This ongoing collaboration started in large parts due to their pedagogical planning, their continuous use of technology in the classrooms and their extensive work with different learning styles (Dunn & Dunn, 1992). The work presented here falls within the area of so called kinesthetic learning, which is about how we understand and learn through bodily interactions and through moving in the world. The learning styles perspective starts out from the idea that different persons have different preferences with respect to how they engage in pedagogical activities. Out of the different learning styles adopted by the school, the teachers were most challenged to find ways of supporting the kinesthetic learners. A person characterized as primarily "kinesthetic" learn best by incorporating their entire body and movement in the learning process and these persons often have problems sitting still in one place and might prefer dramatizing and learning by doing or acting out. This acknowledgement of the importance of the tactile and kinesthetic dimensions fits well with recent research on tangible user interfaces and the emerging bodily interaction field as described above. In a previous study with the school we explored how learning situations can be created in mixed digital and physical settings, and how the properties of the respective media can be utilized to support exploration and learning (Ahmet et al., 2011). With the Weather Gods and Fruit Kids game, we continue this work by exploring how abstract notions of energy and energy consumption can be turned into something that can be experienced and interacted with in a physical manner using your body and movement.

The Weather Gods and Fruit Kids Game

The game Weather Gods and Fruit Kids is a result of a design process involving three previous workshops, one involving school teachers, and two where the children were introduced to the concept of energy in various ways.

The purpose of the first workshop was to find a concept for the children to explore through kinesthetic learning. Three teachers together with two members in the project group had a brainstorming session around some preliminary ideas. The teachers would reflect on suggestions for each of the ideas, and how they could be incorporated into their own respective teaching subjects where they themselves had difficulties incorporating kinesthetic dimensions. After numerous suggestions, the concept of consumption, charging and preservation of energy through using bodily movement was chosen.

The second workshop was designed as a bodystorming session with 20 children, to explore alternative pedagogical activities involving motion sensors and energy consumption. The children wore plastic stickers attached on arms and legs (see Figure 1), simulating sensors that would detect their energy consumption. The estimated energy consumption was then visualized on a computer screen and was reflected on during and/or after each exercise. The exercises involved throwing a large and small ball where the children would observe and compare the visualization of their respective energy consumption levels, moving to mimic a set of pictures (like a kite, snake or basket ball) and moving according to specific movement types, such as slow motion, robotic movement and big and round movements. The aim of the workshop was to simulate, rather ambiguously but in a simple manner, what it means to consume energy. Through different types of movements involving different body parts with varied intensity and strength, the children would at times comment on each other's movements and how it could be that certain movement types required more energy and why energy levels could at times differ from their own expectations.

In the third workshop the children were equipped with real sensors that would actually estimate their energy consumption. Ten children in groups of two were equipped with Nintendo Wii Remotes, nicknamed Wii Motes, on one arm and one leg. While performing a number of exercises they were encouraged to hypothesize and think about the outcomes from the exercises they were undertaking: i) running vs. walking where the children would firstly eat a slice of apple and then either run or walk while observing on a computer screen the type of activity that "consumed" the eaten apple the fastest, ii) the children would operate a lamp and a fan, while observing which appliance would run the longest on the approximate energy amount found in the slice of apple they had previously eaten, iii) moving to charge a 'smart battery' while listening to audio and vibration as feedback, the children would try to find the movement patterns that triggered a rhythmic sound from the Wii Motes. In this workshop, as well as the previous one, the use of sensors, simple visualizations and exercises were not meant to provide the perfect setting nor the correct procedures for analyzing e.g. energy consumption in real life but were designed to spur discussions with the children about the subject energy.

Figure 1. Workshop 2 (left) with Fake Sensor and Wizard of Oz Prototype. Workshop 3 (right) Attaching Wii Motes to Arms and Legs to Detect Movement and Energy Consumption.

In the two workshops with the children, we found that although the children were exploring consumption and preservation of energy, they did it in a very controlled and structured manner. The activities did not contain much free exploration and experimentation around the concepts. The most striking result from these workshops was however how much the presence of a computer screen affected and restricted the movements in the room as well as the interaction between the children. Based on these findings we decided to continue with the relation between body and energy, but in a more self-contained activity were the technology would restrict the activity in the room as little as possible. From these requirements we ended up with an idea of a game activity without computer screens taking place in a large space like a gym hall.

The Game

The Weather Gods and Fruit Kids game engage four players to compete against each other in pairs by physically moving about in a large space, a gym hall, and simulating the generation, consumption and preservation of energy (see Figure 2). All players wear Wii Motes attached to one arm and one leg. The objective of the two fruit kids is to collect fruit cards that are spread out around an obstacle course, while not touching the ground. The motion sensors on the arms and legs of the players are used to estimate the amount of energy consumed when moving around the track. If not moving carefully, the players will run out of energy and thus lose the game. By collecting fruit and bringing it to their 'nest' they will refill their energy levels.

The opposite team, the weather gods, tries to obstruct the fruit kids by stealing their energy. By engaging in various magic dances, they charge energy that can when accumulated be released in the form of thunder and lightning (by pressing a button on the arm-mounted Wii Mote), which will reduce the energy level of the fruit kids. Moving around on a stage placed next to the obstacle course, the gods have a good view of where the fruit kids are and their actions. Rather than performing specific gestures or steps the weather gods interaction is designed for free flowing movement of a specific type: *Big & Fast*, *Slow-motion* and *Robotic*. In *Big & Fast* mode the movement is free but needs to be large and quick, in *Slow-motion* the movements need to be slow, smooth and continuous while in *Robotic* the motion should be very sharp, edgy and staccato. If the fruit kids manage to collect all the fruit without running out of energy they will replace the weather gods, and can be challenged by a new team of fruit kids.

The design was inspired by three of the interactional qualities introduced by Tholander & Johansson for *whole body interaction*: connecting to the physical space, harmonizing modalities and open-ended response (Tholander & Johansson, 2010). To allow for a head-up interaction (Soute et al., 2009), giving the player the possibility to freely explore movements in the physical space and be part of and aware of the larger activity in the hall, the interaction was designed specifically to not occupy the player's full attention, e. g. by visual feedback on a screen. Instead, to allow for a continuous attention to the activity and each other, harmonizing modalities was chosen in the form of vibration and audio feedback to support their understanding of their status in the game. The Nintendo Wii technology was used for motion sensing as well as for tactile vibration feedback both for weather gods and the fruit kids. The current energy level of a fruit kid is represented as individual feedback through the use of vibration pulses with varying frequency; high energy - high frequency vibration and vice versa. For the weather gods the response was designed to be semi open-ended to allow for individual interpretation by the player, and no specific instructions were given on how to know when enough energy had been accumulated. The vibration would indicate whether the user's current body motion was the expected one, by vibrating at a regular time interval. Consequently, a lack of vibration could alert the user that the right dance style is not being danced.

Different forms of audio feedback was also used in the game, e.g. to inform when the fruit kids were out of energy, when lightning was thrown or when fruit cards were collected and new energy had been charged. All these sounds were provided through shared speakers that all players would hear. Individual sound feedback was also provided to the weather gods through the Wii Motes on the arms to indicate whether they were performing the indicated movement type. These sounds were very faint and were mostly only perceived by the bearer of the Wii Mote. The only visual feedback provided by the system was through a light flash that would light up the room whenever lightning was thrown by the weather gods. The researchers played no active role in the game other than indicating a change of dance style to the weather gods by holding up large signs from the tech booth situated inside a smaller room adjacent to the game area.

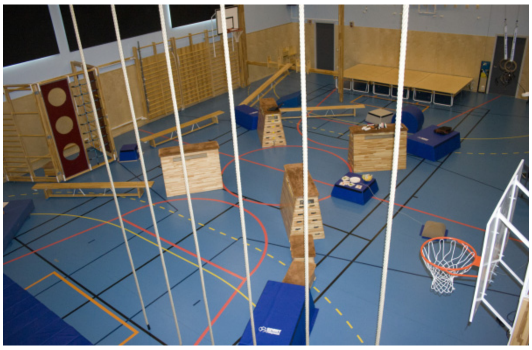


Figure 2. The game field with fruit kids' obstacle course and the weather gods' stage created in a gym hall.

Study and Method

Eleven children between the ages of 9 to 11 participated in the study. The study began with a sum up of the two previous workshops, an introduction to the game and a warm up session with playfully moving about according to the three weather god dances. Four children at a time would play the game while the others were observing on the side. A total of six game sessions were played followed by semi-structured interviews with questions formulated to gather qualitative data on how the activity was perceived, the purpose and also of the understanding of the underlying discussions around the concept of energy. The interviews and the game sessions were video recorded, with one camera filming an overview of the hall and the second filming the stage.

Findings

The particular focus of the findings presented here is on the different ways children use their bodies in the interaction, the kind of experiences they have, and the role the technology plays in these processes. The analysis is framed around the three themes presented by Jonsson et al. on design of spaces for embodied performance. The themes were adapted to the analysis of the interaction and learning in and around the Weather Gods and Fruit Kids game. The themes were interpreted and viewed upon in the following way; *Embodied performance* refers to how the activities take place in an open space without computer screens as dominating source of interaction with the participants. We also focused on the communicative function of different actions in the game, of how the actions performed by the actors in the game were not only directed towards the system and/or the game, but also played a performative role, directed more towards the other players and the spectators. *The role of the physical space* refers to how the configuration of the physical setting affects and restricts the interaction in numerous ways. The role of the physical space becomes especially interesting in this case since the interactive system is situated in a gym hall, which not only provided a large space for interaction and performance but also a space where the children are accustomed to moving and bodily ways of playing. *Device specific qualities* refer to how the properties of interaction devices and objects shape how the overall system and the game as a whole is perceived and interacted with. A particularly prominent aspect was how the setting and the devices encouraged the children to put significant efforts into interacting with the partner, the surrounding and the co-players rather than the technology. The findings are presented below, divided into each theme.

Embodied Performance and Experience

When performing the different weather god dances, the children were observed moving their bodies and different body parts equally much, with no specific emphasis on the arm or leg with the Wii Motes attached. Although instructed that the Wii Motes would sense and pick up their movement, they did not move that arm or leg more than the other; instead their whole bodies were involved and engaged. In a sense they were doing more than needed of movement, with the focus on the dance rather than the Wii Mote. In most cases with the weather gods the Wii Motes were just there, not dictating their movement or focus but was rather an extension of the body, creating an embodied experience and augmenting the child's ability to interact in the game. The children had done the same movement types in the previous third workshop, but had then moved in a more controlled and smaller manner. In the game however, when performing the very same movement types, they did so in a much more elaborate and free way. The game setting and its context, with the Wii Motes and the beforehand decided dances, seemed to give the children a frame within which it was allowed to play out and move in strange ways. While waiting for their game to start, one pair of weather gods would try out the different movement types and mimic shooting out a 'flash' by aiming with their arms out towards the fruit kids tracks. One pair pretended to be hit by each other's flashes by shaking their bodies and laughing. In this *warm-up* some couples would also rehearse movements to do later on in the game, like one pair doing a skier-movement. The freedom of how to move brought out the element of *choreography* in how many weather gods came up with their own movements.

While on stage, some weather gods were seen making weird, tricky and complicated movements involving the whole body from snaking with the head, twirling their fingers around and letting the experience be seen also in their facial expressions (see Figure 3). One pair doing the robotic dance made 'sch-sch' sounds to go with their movement, when trying to get into a rhythm. The freedom space in the game context seemed to allow for the children to tailor their own individual experience of really getting into the movement. This embodied experience was also shown in how many gestures and actions were *exaggerated*, for example the sending off of a 'flash'. The action was described as pushing a button on the Wii Mote, but on stage during the game the weather gods were seen to be in a more embodied state with not only pressing the button, but instead using their whole body to aim with the arm fully stretched out forward from the shoulder. Some were even squinting with their eyes as if looking through a riflescope to aim the flash (see Figure 3).

Figure 3. (Left and middle) Weather gods getting into the grove of moving, acting out with the whole body, and (right) aiming with the whole body to shoot a flash.

The Role of the Physical Space

In the whole body interaction in large space as a gym hall, the interaction space could be seen as divided into personal, near and far space. Firstly the weather gods would interact with their Wii Motes within their own personal space, their bodily sphere. One weather god moving in Robotic style was constantly changing the gaze of his eyes with sharp movement with the head to different directions showing how the interaction allowed for *zooming out* from what was happening in the surroundings and opened up for *zooming in* on the own personal experience and expression of the movement.

The second interaction space, the *near space*, included the fellow weather god on the stage. Despite the relative large open space, the weather gods tended to move around *close to each other*. When interacting in *body-to-body* style, i.e. doing fake hitting movements at each other in slow-motion, the closeness was natural but was also observed when moving freely but following each other around. When moving this close to each other they would also at times *affect the movement space* of the other person and hence also this persons movements (see Figure 4). E.g. one pair moving close by each other in slow motion was seen to adjust their bodies around each other with one moving her torso to make room for the other persons arm to slice through that personal space.

Lastly the stage was part of the larger gym hall space, and situated very close to the fruit kids. The children on stage would mostly have their focus on each other or on their own movement. Sometimes however, events would make them become *aware of the far-space*. For example at one point the weather gods were looking only at each other but when hearing the loud 'eat'-sound, meaning a fruit kid had collected a fruit card and had raised his/her energy level, they looked up at the fruit kids. Then one of the weather god girls tapped the other one of the shoulder, said something in a low voice where after they both started moving more intensely to generate more energy. Many weather god pairs also talked about waiting for the fruit kids to get new energy from fruit cards before they sent off a flash, showing how the interaction allowed them to stay aware of the surrounding environment, what was happening out on the fruit kids track, despite their at times deep dive into the movement and their personal sphere.

Figure 4. Two weather gods influencing each other's movement space when doing a fake hitting action in the slow-motion dance.

Similar behavior was also seen with the fruit kids where in most situations they had their eyes on their own search for fruit cards. However, a sudden flash thrown by the weather gods made one fruit kid pair look up, where after they both started moving at a quicker pace (see Figure 5). They would jump down and almost run into their nest instead of calmly moving about as when they started the game. Despite the fact that they knew that moving in a faster manner will make them loose more energy, it was over ridden by the time pressure of getting their fruit cards to raise their energy level. In another session both fruit kids started to hurry up their pace after throwing glances at the stage where the weather gods were getting ready to fire off a flash.

Another example of becoming aware of far space happened when one weather god pair changed dance style into *Big & Fast* and started their beforehand rehearsed skier-movement, suddenly making a lot of thuds and noise on stage. One fruit kid on a path very close to the stage then started to move in a more hurried way. Prior to this, she had moved in a controlled and slow manner, to conserve her energy, but the *loudness of movement* from the stage made her look up and then move faster (see Figure 5). For the fruit kids, the importance of their role was to find a balance between movement and pace, in order to balance consumption and preservation of energy. While they were focused on finding and retrieving fruit cards to their nest, they would consider the implications of moving too fast (naturally leading to them consuming energy) and moving too slow (and becoming the targets of the weather gods' lightning).

Generally throughout the sessions both weather gods and fruit kids focus was observed to be mainly on the personal and near space but how the children spoke of their opponents in the game also shows how they did have a connection to the environment, that the interaction with the technology did not hijack their focus from being aware of their surroundings. Sound announcements from the speakers as well as noise from large movements on stage, did also alert the players to look up and observe the other team.

Figure 5. (Left and middle) Fruit kids keeping an eye on the weather gods before and while a flash, and (right) fruit kid becoming aware of the game situation by the loudness of movement from the weather gods.

Device Specific Qualities of Interaction

The vibration from the Wii Motes to the weather god's movement was an *open-ended response*, and no instructions were given as to when and how it would feel when enough energy was loaded to send off a flash. When asked how they knew that they had enough energy to send off a flash, one weather god said: *"When it had a very high pulse",* showing with one hand over the wrist where the Wii Mote was attached. Another God said that when "*it was shaking much*" they would alert each other through a code-word to coordinate to send off a flash. Sometimes the technology was not working properly, with the Wii Motes not being active, but those weather gods would find interesting ways to work around the tech gap, with one child saying *"I didn't feel* *anything from mine, but I heard that Otto's was going on like this"* as to how he would hear the vibrations of his partner's Wii Mote as a cue for sending off a flash. Even though this was a discrete sound it apparently caught the attention despite the surrounding noise. This *personal feedback became a joint feedback* in between the weather gods and helped them coordinate their actions.

For fruit kids it was rather *the absence of vibration* that caught their attention. When asked what was a bit challenging in gameplay, they said *"One felt when there was little left, and if one wasn't back in the nest then I thought* 'I have to hurry' *and started moving faster but then I thought* 'no, I have to take it slow otherwise…*"*, showing how also the absence or the very weak vibration feedback made them notice and become aware of their status in the game. The low volume sound emitted from the The Wii Motes of the weather gods was an open-ended response. One child said *"And then there were these tiny clicking sounds, when one was doing the right movement"* about how he interpreted it as doing the right movement type. This was a rather faint sound (as the Wii Mote speakers did not allow a strong sound) and thus became more of a personal feedback.

When basing our interaction on the familiar Wii technology we were half expecting to see some children just shaking the Wii Motes to generate energy instead of moving according to dance styles. Contrary to this though, the children were in most cases moving as if the Wii Motes were not there, e. g. when acting as a robot and mimicking it with their whole bodies. Those times when it did come into focus were situations when the technology did not work or the Wii Motes fell out of the attachments to the body. Also the participants seemed to be aware of the Wii Motes more in the very beginning of their game. One fruit kid began to hold the arm where the Wii Mote was attached rather stiff through the track, but as the game progressed the movements became more smooth and he would move freely also with the Wii Mote arm. Even the weather gods would check that they were securely fastened but when the game progressed they would move more as one entity with the Wii Mote, rather than focus on it. From the weather gods there was a natural focus on our tech booth as they would receive cues for changing dance style by us holding up a sign. Yet they would aim to shoot a flash at the fruit kids and not towards our tech booth where the loudspeaker and the flashing lamp was situated.

In the interviews, the experience of the vibration feedback on the skin did mostly arise as a topic when describing different peak situations, such as almost being out of energy or being ready to fire off a flash. Fruit kid players did not speak specifically of the healthy high pulse that meant they were doing fine and had energy to take them through the tracks, but talked about being aware of it when the pulse grew weak and the energy started to run out. The loud audio feedback from the speakers would make the children look up and make them aware of changes in the game. When waiting in line to play, the children would observe the games and as the session progressed they learned tactics of how to move to win. But when asked what they understood when they were observing, one boy said *"one didn't know how it felt"* and his friend *"and how fast one could move before it ran out*", in support of how the vibration feedback actually did help them understand their status in the game.

Discussion and Conclusions

The gym hall setting for the activity and our choice of using several different modalities had the effect that there were no obvious points of focus for shared information about the status of the game. Even though based on and dependent on technology, the activity became to a certain extent de-coupled from the computer system. Instead of focusing on the technology, the different groups had their focus on their immediate activity and movement and would from time to time update themselves by looking around or be made aware of the overall status in the game by audio cues and events happening out in the hall. The system provided feedback and information in multiple modalities: personal feedback in terms of on the skin vibration, joint sound feedback in the near space, and public feedback through audio and light. The effect of designing open-ended vibration and sound feedback yielded personal interpretation and enabled an individual experience. The vibration feedback proved to be a strong indicator and was most noticeable when situations reached some kind of emergency limit, e.g. when energy ran low and the pulse grew weak, suggesting that the healthy pulse remains in the periphery until its absence is discovered. It is felt via the body, just as we are unaware of the states of our body until a change occurs that make us consciously focus on the body part that signals the change. This suggests how feedback felt by the body could be powerful to use for interaction in environments where the user's attention is highly distributed. Specifically, it highlights how the absence of a feedback to the body can be used as a strong response to catch a user's attention.

The game allowed for an element of performance in how the children would do more than needed by the technology sensing their movement. In contrast to an earlier study of bodily interaction with a digital artefact where the players spoke of trying to please the technology they were interacting with (Tholander & Johansson, 2010), the children playing this game did not move just because the Wii Mote needed it but because their experiences was built from movement and embodied interaction. It would have worked equally well to interact with the system by only moving the Wii Mote arm and leg, but instead all children, moved their whole bodies in rich ways. How the weather gods children would move in free, creative and engaged ways adding facial gestures and sound effects, showed how the context of the game allowed for personal exploration a space of freedom to tailor the activity to their own preference. These weird, exaggerated movements performed for the pleasure of moving, shows how the activity opened up for an embodied experience with open-endedness in interaction and response. It is also noticeable that for bodily movement interaction, the 'safe zone' or frame for moving created by the game, opened up the creativity space for the children to dare to explore using movement.

The large space gym hall setting for the activity gave different dimensions for experience, from embodied interaction in the personal space to being part of the larger whole in the hall. To successfully interchange between near and far space, the interaction should make use of both system response (like audio) but also alerts from the real world setting (of thuds, noises, or other naturally occurring sound from other movers in the surrounding). We argue that by adapting the interchange needed between the different spaces, it is possible to build activities and learning environments that allow for both a rich experience of small details in the personal space to a joint understanding of concepts in the larger whole, through both system and real world feedback and interaction.

The aim of the project was to explore the possibilities to spur discussions in a subject through designing for learning by using bodily movement in a large space setting. Based on the outcomes of our design process and the game, we argue that this shows a promising approach for designing for alternative pedagogical practices that supports in particular kinesthetic learners in their need to learn by being physically engaged.

References

Ahmet, Z., Jonsson, M., Sumon, S. I. and Holmquist, L. E. (2011) Supporting embodied exploration of physical concepts in mixed digital and physical interactive settings. TEI '11. ACM

Dunn, R. and Dunn, K. Teaching elementary students through their individual learning styles. Allyn & Bacon Inc, Boston, 1992.

Hornecker, E. and J. Buur (2006). Getting a Grip on Tangible Interaction: A Framework on Physical Space and Social Interaction. CHI 2006, Montreal, Canada, ACM Press.

Höök, K. (2010). Transferring Qualities from Horseback Riding to Design. In Proceedings of NordiCHI, Reykjavik, Iceland, October 18 - 20, ACM Press

Jonsson, M., Tholander, J., Fernaeus, Y. (2009). Setting the stage – Embodied and spatial dimensions in emerging programming practices. Interacting with Computers 21(1-2): 117-124.

Klemmer, S. R., Hartmann, B., Takayama, L. (2006). How bodies matter: five themes for interaction design. DIS '06, ACM Press.

Laaksolahti, J., Tholander, J., Lundén, M., Solsona, J., Karlsson, A. (2011), The Lega: A Device for Leaving and Finding Tactile Traces, Proceedings of TEI' 11

Lakoff, G., Johnson, M. (1999). Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought. New York, Basic Books.

Larssen, A.T., Robertson, T., & Edwards, J. The feel dimension of technology interaction: exploring tangibles through movement and touch. TEI'07, (2007), 271-278.

Moen, J. (2005). Towards People Based Movement Interaction and KinAesthetic Interaction Experiences. In proc. of CC '05 decennial conference on Critical computing, August 21-25, Aarhus, Denmark.

Price, S., Sheridan J. G.et al. (2010). Action and representation in tangible systems: implications for design of learning interactions. Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction. Cambridge, Massachusetts, USA, ACM: 145-152.

Schiphorst, T. soft(n): toward a somaesthetics of touch. Ext. Abs. CHI'09, (2009), 2427-2438.

Soute, I., Markopoulos, P., Magielse, R. (2009). Head Up Games: combining the best of both worlds by merging traditional and digital play. Personal Ubiquit. Computing, SpringerLink

Sundström, P., Ståhl, A. and Höök, K. (2007). In situ informants exploring an emotional mobile messaging system in their everyday practice. Int. J. Hum.-Comput.Stud., 65 (4). 388-403.

Säljö, R. (2000). Lärande i praktiken - ett sociokulturellt perspektiv. Bokförlaget Prisma.

Tholander, J., Johnssson, C. (2010). Bodies, boards, clubs and bugs: A study of bodily engaging artifacts. CHI10' Extended Abstract on Human factors, April 10-15, Atlanta, GA, USA

Tholander, J., Johansson, C. (2010). Design qualities for Whole Body Interaction – Learning from Golf, Skateboarding and BodyBugging. NordiCHI'10, ACM Press

Ullmer, B. & Ishii, H.(2000). Emerging frameworks for tangible user interfaces. IBM Systems Journal, 39(3-4), 915-931.

Zangouei, F., Babazadeh Gashti, M. A., Höök, K., Tijs, T., de Vries, G-J., Westerink, J. (2010). How to Stay in the Emotional Rollercoaster: Lessons Learnt from Designing EmRoll. NordiCHI '10. ACM Press

Ziegelbaum, J., Millner, A., Desai, B., Ishii, H. (2006). BodyBeats: Whole-Body, Musical Interfaces for Children. CHI 06' April 22-27, Montréal, Québec, Canada, ACM press

Acknowledgments

We would like to thank the children and teachers at the elementary school for their ideas, inspiration and hours of fun and creative participation. Also thanks to our colleagues for their input and comments.

Improving the mCSCL Approach of a Mobile Chinese Character Forming Game via a Design-based Research Cycle

Lung-Hsiang Wong, Learning Sciences Lab., National Institute of Education, 1, Nanyang Walk, Singapore 637616, lunghsiang.wong@nie.edu.sg
Ivica Boticki, Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, Zagreb, Croatia, ivica.boticki@fer.hr
Jizhen Sun, Chinese Culture University, 231 Section 2, Jian-Guo S. Road, Taipei, Taiwan, ccsun@sce.pccu.edu.tw
Chee-Kit Looi, Learning Sciences Lab., National Institute of Education, 1, Nanyang Walk, Singapore 637616, cheekit.looi@nie.edu.sg

Abstract: We describe one cycle of design-based research (DBR) in which we explore mCSCL through an iterative process of (re)designing and testing the learning approach with students. The mCSCL application assigns each student a component of a Chinese character and requires them to spontaneously form groups that can assemble a Chinese character. We observe the enactment of the learning design in two modes (with and without the ICT), and found the students favor the card mode over the phone mode due to their emergent trial-and-error strategy. That triggered us to examine the scaffolding strategies by exploring domain-oriented theories to inform us in deciding how we should accommodate their use of the strategy. This cycle of DBR has reshaped our learning design. This paper brings to the fore the value of the interplay of theories, implementations and reflections as advocated by DBR.

Introduction

The proliferation of networked computers supplemented with software applications has provided us with a new network-based medium to design educational services. A variety of studies in the field of mCSCL (mobile Computer Supported Collaborative Learning) has explored opportunities for designing learning applications through mobile technologies (Looi, Wong, & Song, in-press; Zurita & Nussbaum, 2004). While such innovations coupled with learning design look promising, we need to deal with the challenge of adoptability in real classrooms. We adopted a design-based research (DBR) perspective to provide a cyclic process of designing, testing and redesigning the learning design and applications, and to integrate design principles with technological affordances to render plausible solutions. Our goal is to conduct reflective inquiry to test and refine innovative learning environments as well as to refine new learning design principles.

Building on prior studies, we propose the design of a model for a mobile synchronous collaborative learning game with the unique characteristic of spontaneous small group formations. In the game, students follow or adapt the collaboration rules or scaffolds imposed both by the teacher and by the computer system. To complete the learning tasks, they must draw upon their social relationships with other students to negotiate acceptable solutions. Previously, we developed a fraction addition game system and did trials with primary school students. This work yielded positive findings in students' emergent collaborative strategies (Boticki, Looi, & Wong, 2010; Boticki, Wong, & Looi, 2010). We then reused the generic software architecture and game model to implement "Chinese-PP", a game-based learning approach on collaborative Chinese character formation. PP refers to 拼一拼 or "Pin yi Pin" in Chinese, which roughly means "trial assembling". Our DBR methodology allowed us to collect and analyze data to many factors simultaneously and to use the rich data to iteratively improve a design more rapidly than might be accomplished through systematic experimentation on each individual factor (Design-Based Research Collective, 2003). Echoing Roschelle et al.'s (2010) call for transforming handheld collaborative tools to classroom modules, our ultimate goal is to elevate our game model to a pedagogically-oriented learning environment that can facilitate the students' learning growth over the time, rather than drilling them through repetitive game playing. We envisage a series of learning activities to be conducted on a regular basis with varied scaffolds and rules undergirded with a theoretical framework.

Now that we have the mCSCL tool, our discovery of the students' collaborative patterns in the fractions game and a theoretical framework for pedagogical design, we would like to adapt the model and the tool to a different domain – the forming of a Chinese character from components. The students involved in Chinese-PP game may demonstrate different collaborative patterns and encounter different challenges in achieving their game tasks as compared with the previous fractions game. Therefore, it was necessary for us to conduct experiments with students using Chinese-PP as part of the DBR process of building the learning environment.

This paper focuses on such a pilot trial of Chinese-PP, which was conducted in two modes, namely, one using digital technology - the "phone mode" and the other without the digital technology - the "card mode". The trial enabled us to examine the domain-independent and –dependent collaborative patterns and attitudes of the students in each mode. With that, we gained a better understanding in the affordances of mobile technologies in facilitating such a game. The findings will inform us in the subsequent revision of both the game model and the mobile application, as well as assist us in fine-tuning of our pedagogical framework.

Background Literature

Chinese Character Learning

The Chinese script has always been the biggest challenge for learners of Chinese (Wong, Chai, & Gao, 2010; Y. Zhu & Hong, 2005). Fan, Tong and Song (1987) claim that the logographic nature of the Chinese script constitutes the primary hurdle. Shen (2002) attributes the challenge to the retention of the combination of sound, shape and meaning of a character in learners' long-term memory. This makes it even more difficult for a learner whose first language is based on an alphabetic writing system to master Chinese (Ho, Ng, & Ng, 2003).

The Chinese scripts are a principled and rule-based system – each Chinese character comprises of one or more components, spatially arranged with certain principles (Liang, 2004). Most of the components have fixed roles to play, as either a semantic component or a phonetic component; only a few of them play both roles. Furthermore, the number of commonly used characters for learners (1,000-3,000) is much larger than the number of component types (< 100). The number of characters' spatial configuration is also limited and rule-based. Zhao and Jiang (2006) propose that there are 10 basic spatial configurations for characters (see Figure 1).

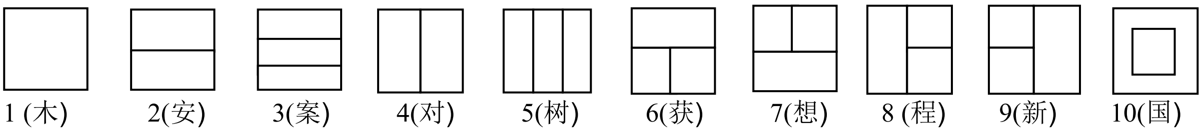


Figure 1. 10 Basic Spatial Configurations for Chinese Characters (with One Character Example Each).

Studies (e.g., Wang, Perfetti, & Liu, 2003; Z. Zhu, 2004) have indicated that those who have acquired Chinese characters recognize them mainly based on their structural elements such as graphic forms and spatial configuration, treating each character as a salient perceptual unit. Tan and Peng (1991) also argue that analyzing the 3-dimensional characteristics (spatial configuration, semantic element and graphic form) is the necessary route leading to the effective recognition and reading of characters. Informed by language acquisition theories (e.g., Comprehensible Input (Krashen, 1985), Information Processing (Bialystok, 1978), and Connectionism (Gasser, 1990)) and Bloom's Taxonomy, we argue that there are six steps in acquiring Chinese characters, in hierarchical order: comprehension, combination, memorizing, application, analyzing, and creation. The fact that a limited number of semantic components and phonetic components can form a large number of characters leads us to argue that learning characters through rearranging and combining their components in different positions is cognitively effective, as it allows learners to comprehend, remember and apply the principles of character formation. This process also has the potential to nurture the ability of using educated guesses when they encounter unfamiliar characters in reading (e.g., a character with the component 氵 is likely to carry a meaning relevant to liquid, e.g., 河 = river, 湿 = wet).

Mobile-assisted Language Learning and mCSCL

In recent years, we have witnessed the paradigmatic development of the Mobile Assisted Language Learning (MALL) framework in enhancing language learning (Wong, Chin, Tan, & Liu, 2010; Wong & Looi, 2010). The focus of MALL research is shifting from content-based (delivery of relatively static learning content through mobile devices) to design-oriented (design of authentic or social mobile learning activities) studies (Kukulska-Hulme & Shield, 2007). In particular, Zurita and Nussbaum (2004) tapped on Syllable-mCSCL, a 1:1 (one-device-per-student) mobile learning game to facilitate Spanish vocabulary learning for young children. In their game design, the students were given language tasks that they had to solve by working in groups of three. A syllable is assigned by the system to each group member's mobile device (e.g., "si", "la" and "bi") and the three students need to determine sequences of the syllables to form correct Spanish words (e.g., "silabi").

Latest developments in the field of mCSCL extend the idea of mobile learning with the collaborative scaffolding in order to include both social and epistemic collaboration scripts encouraging small group participation (Nussbaum, et al., 2009). The design of collaborative scaffolding should encourage social interactions, facilitate joint problem solving, lead to richer knowledge construction and at the same time take into account different and emerging roles, joint group goals and actions and facilitate verbal explanations. These developments mesh well with our goal of designing a component-based Chinese character learning activity through socio-cognitivist and socio-constructivist means.

Towards a One-to-One mCSCL Solution for Chinese Character Learning

Building on the work of other researchers (Zurita & Nussbaum, 2004), we developed the Form-A-One (FAO) System, a mobile collaborative learning game with flexible, small group sizes. In Nussbaum's work, students are assigned to fixed small groups before the beginning of the activities. We adapted Nussbaum's design so that students would have to find and negotiate with other students to form their own groups spontaneously with no fixed size. The activity is conducted in multiple rounds. In each round, a set of fractions are assigned by the system server via 3G connections to individual students for them to form groups with the sum of the members' fractions of each group equals one (e.g., a group of three with 3/6, 1/4, 2/8). When the game advances to the next round, the existing groups are disbanded and a new set of fractions is assigned to the students. The activity consists of three main scaffolding sources: technological, social and the teacher. Technology provides scaffolding in both generic and context-specific rules and logic (in the form of software affordances), while the teacher acts as facilitator and helps the students in dealing with impasses. Social scaffolding is encouraged in order to increase student interaction and collaboration (Boticki, Looi & Wong, 2010).

We conducted a pilot study in late 2009 that involved 16 Primary 3 students (Boticki, Looi, et al., 2010) in Nan Chiau Primary School, Singapore. One important finding was the students' modification of their initially chosen ad-hoc strategies (e.g., gender or personal preferences, looking for the same fractions, randomly sending out invitations, etc., which inevitably ended with impasses) coming as a consequence of them realizing the importance of achieving the global goals besides their local group goal, therefore learning how to collaborate (e.g., breaking out groups for improved solutions). We then analyzed the process by examining the evidence and what it took, in the form of teacher, technology or social scaffolding, to achieve the four elements/principles of cooperative learning (Johnson & Johnson, 1994): positive interdependence, maximum peer interaction, equal opportunity to participate, and individual accountability.

Similar to the fractions game, the Chinese-PP activity is conducted in multiple rounds. In each round, a set of Chinese components are assigned by the FAO system server to individual students' smart-phones (with the Chinese-PP client application installed – see the screenshots in Figure 2a). Students are required to form groups by choosing appropriate characters out of the assigned components, thereby forming a valid Chinese character. Members of each group then discuss and choose one of the general Chinese character configurations (see Figure 1) to organise their components properly via templates (character configurations) supplied by the Chinese-PP application (arrows < and > in Figure 2a). For example, with the components覀, 示 and 风, students could decide to choose template no. 9 (Figure 1) and place the components in the correct order to form the character.

Figure 2a. A student's smart-phone application showing an assembled character out of individual components.

Figure 2b. A student's smart-phone application showing a list of student with components participating in the Chinese learning activity.

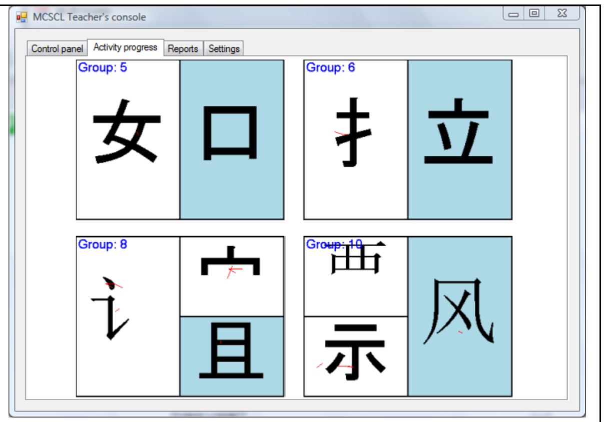


Figure 2c. A teacher's view of the Chinese language learning activity (the "Group Status Window"). The screen shows the characters from four groups of students. Each group has an assembled character framed into a template chosen by the students on the device.

In preparing each round of the game, the teacher needs to select a set of components according to the number of participating students and input them to the system. The choice of components should allow the construction of as many eligible characters as possible, and with at least one global solution (i.e., no component/student will be left out) available. For example, for an eight-player game, a possible component set is [木 又 寸 宀 女 禾 口 王], where students could form three groups and construct the characters [树 安 程] or [案 对 程] without any player being left out. However, there exist other combinations such as [宋 对 和], with王 and女being left out (no character with the combination of these components), and a lot more. Although students should be encouraged to socially figure out a combination where all the components are used to form characters, it is a tall order for our target students to achieve that, given their limited language proficiency (Chinese L2 standard) and cognitive ability (Primary school level). However, we had been keen to introduce game rules to motivate them to form characters with as many components as possible and try to minimize the number of left-out peers. In turn, students who have formed groups should continue to explore other possibilities, perhaps by inviting another peer to join the group and form a new character (e.g., a group with the character 宋may invite their peer with the component女to form案), or even disband an existing group, combine and reshuffle with other components to form two or more new groups!

Micro-cycle of Design, Evaluation and Re-design with Chinese-PP

We ran the first implementation of Chinese-PP in Nan Chiau Primary School, involving 37 Primary 4 (10-year-old) students. These were mixed ability students in Chinese Language and had had one year's experience in using HTC TyTN II smart-phones on a 24x7 basis (and were therefore adept in using the device) for learning science, English and math. Among them, 16 students were also involved in the trial of the FAO fractions game in the previous year. As such a game may also be carried out, for example, using cards with individual character components being printed, we experimented on both the "phone mode" and "card mode" (with four rounds of each game) on two different days. The students were split into two subgroups. Subgroup A with 19 students played one hour of card game followed by one hour of phone game. Subgroup B with 18 students went through both games in reverse order. Table 1 shows the schedule of both subgroups' Chinese-PP experiences.

Table 1: The experimental process of Chinese-PP.

	Day 1		Day 2	
Subgroup A	Card game A-1	Focus group FA-1	Phone game A-2	Focus group FA-2
Subgroup B	Phone game B-1	Focus group FB-1	Card game B-2	Focus group FB-2

The games were played in a special classroom with more open space than usual so that the students could freely move around to negotiate with different peers in group forming. For the phone games, the students could invite potential group members and accept/reject invitations through the phone application. The teacher facilitated all the games by controlling the game pace, hinting the students on-the-fly concerning possible groupings, verifying students' groupings, and determining when to terminate a round.

We video and audio recorded all the games to document and code students' game behaviors and collaborative patterns. The software logs of the students' interactions during the phone games were also used for triangulation. In addition, we conducted focus group interviews as noted in Table 1 (each session involved six students, three of whom had also played the previous fractions game) in order to find out their perceptions in the games and the reasons behind the collaborative behaviors that we observed. Due to the space constraints, we will not show detailed coding and analysis of students' spatial movements and discourses, but will present high-level findings and how they have informed us in our subsequent adaptation of the learning environment.

The reason that we designed the experimental process in this way (Table 1), as opposed to the usual experimental-control group design, is that we wanted to let both subgroups of students to experience the game played in both modes. This approach permitted us to investigate the changes of their perceptions and to see if that might be influenced by the order of experiencing the two modes. Nevertheless, our initial analysis on the focus groups shows that both subgroups of students have expressed fairly consistent perceptions in both modes, and therefore the factor of "which mode was played first" did not have any substantial impact in our analysis.

Findings

Domain-independent Collaborative Patterns

In all the card and phone game sessions, the students exhibited similar discussion patterns as in the previous fractions games. Note that in both types of games, we encouraged the students to form legitimate groups with more peers, and to minimize left-out peers. Typically, in the beginning of a game round, students started exchanging ideas verbally about arranging the components. Most students started with identifying a classmate each to discuss, and then expand from pairs to clusters of 3-4 to discuss alternative possibilities. An initial set of groups was created in the process, with a few left-out students still looking for groupings. However, personal and gender preferences often influenced the formation of these initial groups. If the small amount of left-out students could not form new groups among themselves, they would seek peers' or the teacher's assistance in identifying other solutions. Meanwhile, some students who had already formed groups had considered helping their left-out peers by thinking of the possibilities of adding a group member or even splitting their own group. There came minor changes in the groupings with the effect of personal and gender preferences gradually fading out. Nevertheless, such positive tendencies were not good enough to make the students achieve any global solution (i.e., no peer was left out) in all the Chinese-PP sessions, which we will explicate below.

The gender-shy issue is relatively minor but still shows intriguing influences in the game dynamics – as it might result in sub-optimal solutions at the early stage of each game round which potentially jeopardizes the chance of eventually reaching a global solution. In the card games, as face-to-face (f2f) discussion was the only mode of collaboration, we observed that students almost always started with physical clusters of the same gender to carry out initial explorations. Students who were left out at this stage then proceeded to move around the classroom to seek groupings with members of opposite gender. Conversely, the phone games allow two modes of communication – f2f and phone invitation. We examined both the videos and the software logs and discovered that the same groups of students were less reluctant to invite members of the opposite genders to form groups even in the initial stage. In addition, we compared the final groupings of all the game rounds and found that 65.3% of the groups formed during the card games were comprised of members of the same gender, which was a high contrast with the phone games with only 40.2% of the groups were such. Among the 12 students interviewed during FA-2 and FB-2, 8 of them admitted that they were more gender-shy during the card games but not so during the phone games, while the rest claimed that they were not gender-shy in either mode.

Indeed, with two communication modes being offered by the phone mode, individual students may opt for their most comfortable method in interacting with their peers, which was one of the advantages of the mode. Nevertheless, due to the nature of the domain (Chinese characters) and the software UI design, most of them had surprisingly indicated their preference in the card mode after the study, which we will explicate below.

Domain-specific Collaborative Patterns

During the card games, we observed the students who clustered together often physically manipulated their cards by trial placing them in different spatial configurations. We identify it as a form of social scaffold. Figure 3 illustrates a scenario we observed in one of the card game rounds. Two students who carried the components 女 and 木 respectively tried all four possible configurations to bring them together (i, ii, iii, iv), but none of them formed a legitimate character. Third student who carried 宀 joined them, and tried to match the component with the first two students' components individually (v, vi), and legitimate characters were formed in both cases (宋, 安). They then faced a dilemma of "which friend to 'sacrifice'", thus deciding to try forming a three-component character instead. They formed 桉 (vii) but thought that it was not a character (this is a legitimate but rarely used one that they had not learned before). They eventually figured out vertical placing of the three components and formed the character 案 (viii).

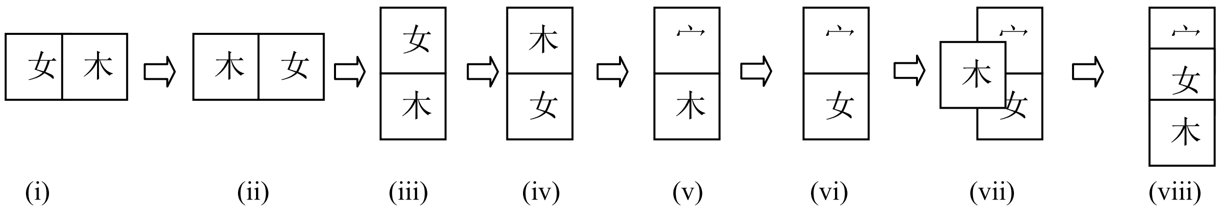


Figure 3. A Scenario of Three Students Manipulating Their Cards in a Card Game.

This is a strategy that we loosely call "trial and error" in this paper. We did not teach them this strategy but they figured it out by themselves. We observed that all students applied this strategy almost all of the time in the card games. The card mode offers the convenience of doing so while the phone mode does not. During the phone games, the students had to study the "peers' components" screen, and mentally construct and picture characters before deciding whom to invite. When playing the phone games, some students still approached each other physically to discuss, albeit not being able to physically "trial construct" characters from their components. In one instance during A-2, two students carried their phones close enough and manipulated their placements to imitate card manipulations (although they found it cumbersome).

At the end of A-2 and B-2, we asked for a quick show of hands to reflect the students' preference between playing card and phone game. More than 80% of them chose the former. The focus group interviews showed a similar tendency, which could be summarized in two points, (1) After A-1 and A-2, most of students had a good expectation in the phone game regardless of whether they had tried it, perhaps due to their technical-inclined mindset after using their phones for a year; (2) Having tried both modes after B-1 and B-2, most students got used to the trial-and-error strategy and found that only the card mode was conducive to that.

Obviously, the UI design and the invitation mechanism of the smart-phone application, both of which were inherited from the fractions game, had been too complicated as well as imposing an additional cognitive burden to them. We asked the students who had played the fractions game before to compare their experiences to the Chinese-PP game. Most of them did not mind playing the fractions game in phone mode as they were capable of the required mental calculations, while spatial configurations were not applicable to fraction additions. This suggests that the application UI was not an issue for this group of students who had been using the phones 24/7 for a year. Instead, it was the domain-specific factor that posed the challenge to them.

Towards the Pursuance of Global Solutions in Game Playing

Regardless of the game modes, the unique game design of Chinese-PP (as well as the fractions game), apart from spontaneous group formation with flexible sizes, is the reinforcement of global solutions (i.e., no student to be left out). In order to work toward this goal, students ought to set aside personal preferences and the pursuance of local (individual) goals to help their peers. While some groups might have formed their characters, the others might have reached a dead-end situation. Students are then required to put their global goal in front of the individual or group goals and start thinking about other possible solutions or group configurations – that is, they need to draw and integrate their thinking and collaborative skills and their linguistic knowledge to achieve this.

In order to encourage the students to work toward the goal, we projected the grouping status window which was originally designed for and only accessible by the teacher, to the students during the phone game. The display offers the up-to-date student groupings where students could study and consider if they could reshuffle or break existing groups to form better solutions. This is a potentially powerful affordance that the card mode cannot offer. However, according to our analysis of the two phone game sessions, the students often took a glance at the display, merely out of curiosity. When we explained to the students during the Focus Group FA-2 and FB-2 on how they could take advantage of the display, they all agreed (including those who did not favor the phone game) that they could benefit from it if they were given another chance to try the phone game.

In addition, after A-1 and B-1, we developed a scoring scheme and applied it in A-2 and B-2. Students accumulated scores by forming legitimate groups – 10 points for a 2-component character (same score to be awarded to each group member), 30 points for a 3-component character, 50 points for a 4-component character, and so on. This was to encourage the students to form bigger groups to fulfil "local" goals. However, in order to motivate them to assist their left-out peers (part of the "global" goal – to reduce the number of left-out students), each student who has formed group and earned points was penalized by 5 points times number of peers left out by the end of a game round. Due to a certain resource and time constraint, we did not automate this functionality in the phone game system. Thus, for both A-2 and B-2, we recorded the scores manually and wrote them on the whiteboard, which was logistically cumbersome and time consuming. However, some students reported to us that the new scoring system did stimulate them to pay more attention in balancing the "local" and "global" goals.

Subsequent Reflections and Further Work

ICT or non-ICT?

The trial seemed to leave us with more uncertainties. Shall we give up the phone mode and settle with the card mode for Chinese-PP due to the nature of the domain? Conversely, what are the potential technological affordances of the phone mode that may justify the mCSCL solution? If the answer is the latter, how can we deal with the perceived "incompatibility" between the phone game and the students' favorite trial-and-error strategy?

We further analyzed the designed and enacted game processes, and concluded that the critical success factor of this game is the intertwining of the local and global goals for the student players that would stimulate their active Chinese character retrievals, thinking of alternatives, applications of component rules, and making educated guesses when encountering unfamiliar characters. It may also further reinforce individual accountability, equal opportunity to participate, positive interdependence, and maximize peer interactions. We believe that the transparency of the global game view (the grouping status window) and the scoring scheme are two potentially powerful scaffold forms which should be retained. Both scaffolds, however, are logistically cumbersome if executed manually. In addition, a minor advantage of the phone mode is that the gender-shy issue could be eliminated at the beginning of each game round, instead of gradually fading out towards the end of a card game. In this regard, we believe that the retention of the technological form of Chinese-PP is justified.

Is "Trial and Error" a Suitable Learning Strategy or a "Bad Learning Habit"?

With that, we decided to provide a theoretical treatment on the emergent trial-and-error method to see if this is a justifiable strategy for Chinese character learning that we should support, or a "bad learning habit" to be rectified – and if that is the case, how can we design scaffolds to help students shake off the habit? We conducted another round of literature review on the relevant issues, which is summarized below.

The acquisition of any subsystems of a language involves how learners establish the connections between forms and their meanings. Most learners attend meaning before form in an attempt to communicate with others (Krashen, 1982). Ellis (1994) argues that the processes for the acquisition of semantic and the formal components of words are distinct. The form-meaning connection is initially made when learners register a form, a meaning and the fact that the form encodes that meaning in some way, or the meaning could be encoded by that form (VanPatten, 2004). Such a process is applicable to Chinese character acquisition as well.

Given the complexity of a form or a meaning that learners should acquire in order to establish the proper relationship between them, it is likely that learners would go through more than one attempt to make the form-meaning connections. Such a view is consistent with Piaget's theory of cognitive developmental stages that include the sensory-motor period, preoperational period, concrete operational period and formal operational period. Piaget suggests that children first learn by actively doing in a random way and, as a result of experience and reconstruction, gradually move to think logically and more objectively during the concrete operational period that spans from 7 to 11 years old (Inhelder & Piaget, 1958). In our experiment, the students' learning strategies reflect the characteristics of the concrete operational period in which they assimilate in a collaborative learning mode using both trial-and-error and logical thinking as the main strategies to reach the next stage.

One of the strengths of mCSCL is its capability for multiple branching and for allowing learners making multiple attempts at problem solving, or in this case, at form-meaning connections. Therefore, we conclude that we should exploit suitable technological affordances to support such a strategy that students find it natural and even fun to apply.

Redesigning the Technological Support

Informed by our reflections, we revamped the technological support for Chinese-PP. A major redesign of the smart-phone application UI will take place. The two key improvements that we envisage are (1) to make the UI more intuitive and convenient for the students to perform their personal and social learning tasks; and (2) to allow multiple learning pathways catering to individual students' preferences (Looi, et al., 2009). We are designing a brand new UI where students will be given a working space on the screen, with their own and other peers' components all displayed at the edges in the form of "virtual cards". The students could easily drag and drop their chosen virtual cards onto and around the working space. The students will not have to construct the characters mentally but can perform a quick personal, virtual "trial and error" before inviting and negotiating with their peers. Furthermore, the invite-accept-reject mechanism will be simplified; an inviter's trial-composed characters will be automatically sent to all their invitees to support the latter's decision making.

Apart from the smart-phone application, the game status window will also be enhanced to support students' global goal pursuance. Other than the formed characters of the committed groups, the window will also display left-out components. Putting all such information in the same display will further stimulate the students to consider alternative groupings that may improve the overall solutions. Score tracking will be automated; and the students' scores and rankings will be updated dynamically in the same window. We hope that the combined use of the personal UI on the phone and the shared display will further reinforce the balancing between the students' pursuance of local and global goals. However, additional teacher scaffolds need to be in place to ensure students making good use of the resources. We will investigate such scaffolds, experiment with them in the subsequent studies, and incorporate them as either technological or social scaffolds.

Transforming Chinese-PP into an Effective Classroom Module

As stated before, we envisage elevating Chinese-PP beyond a one-off or repetitive mCSCL game design to become an ongoing pedagogical practice that facilitates students' language learning growth. We will co-design 8 one-hour learning session with teachers, with varied teacher/social/technological scaffolds and game rules (e.g., a new game rule "allow cloning of components in the same character" such as composing 淋 with two instances of 木, or alerting them on the relevant linguistic rules through the teacher or technological means) across different sessions. We will vary these elements not just for injecting new excitement to sustain the students' interest. We will strive for mapping a theory-informed (language acquisition theories and Bloom's Taxonomy) Chinese character learning process into the variations of the game elements. We intend to conduct a full-fledged study by implementing the new pedagogical model, with the revamped mCSCL infrastructure in place.

Conclusion

We have narrated our journey of conceptualization, prototyping, trial implementation, reflection, and refining of Chinese-PP which is our mCSCL solution for Chinese character learning. We adopted a DBR approach to embark on an iterative process of (re)designing and testing the learning approach by examining the effects of the technological and social scaffolding. In each iteration, we anticipated how the proposed learning design might be realized in a classroom, and how the students might collaborate, interact and learn as they participate in the activities. From the enactment of the learning design in two modes, we analyzed the actual process of student collaboration and learning in each of these two modes. On the basis of this analysis, we drew conclusions and made recommendations for the next learning design which will combine the best of the two modes.

Thus we found versatility and flexibility via the DBR methodology guiding our quest towards a refined solution to address the challenges that we encountered when students used the current design. In prior mCSCL (or learning technology in general) studies, the grounding of domain-specific learning theories tend to be accorded a lower priority. In the early stage of our study, we observed how domain-specific factors undermined the learning model. That is, the students performed well in, and positively perceived the previous phone-based fractions games. However, in the Chinese-PP games, they preferred the card mode due to their emergent trial-and-error strategy. That caused us to further examine the scaffolding strategies by conducting another round of literature review. We let the domain-specific theories guide us in deciding whether we should accommodate or rectify the students' use of their emergent strategy. Our early exploratory study of the project has effectively reshaped our overall learning model design, bringing to the fore the value of the interplay and iterations of theories, implementations and reflections, as advocated by DBR.

References:

Bialystok, E. (1978). A theoretical model of second language learning. *Language Learning, 28*, 81-103.

Boticki, I., Looi, C.-K., & Wong, L.-H. (2010). Doing collaboration and learning fractions with mobile devices. *Proceedings of the Global Chinese Conference on Computers in Education 2010* (pp. 9-15), Singapore.

Boticki, I., Wong, L.-H., & Looi, C.-K. (2010). Can one-to-one computing help children learn cooperatively? *Proceedings of the International Conference on Computers in Education 2010* (pp. 407-411), Putrajaya, Malaysia.

Design-Based Research Collective. (2003). Design-based research: an emerging paradigm for educational inquiry. *Educational Researcher, 32*(1), 5-8.

Ellis, N. (1994). Implicit and explicit language learning - an overview. In N. Ellis (Ed.), *Implicit and Explicit Learning of Languages*. London: Academic Press.

Fan, L., Tong, L., & Song, J. (1987). The characteristics of Chinese language and children's learning to read and write. In D. A. Wagner (Ed.), *The Future of Literacy in a Changing World* (pp. 87-94). New York: Pergamon.

Gasser, M. (1990). Connectionism and the universals of second language acquisition. *Studies in Second Language Acquisition, 12*, 179-199.

Ho, C. S. H., Ng, T. T., & Ng, W. K. (2003). A 'radical' approach to reading development in Chinese: the role of semantic radicals and phonetic radicals. *Literacy Research, 35*(3), 849-878.

Inhelder, B., & Piaget, J. (1958). *The Growth of Logical Thinking from Childhood to Adolescence*. New York: Basic Books.

Johnson, R., & Johnson, D. (1994). An overview of cooperative learning. In R. V. A. N. J. Thousand (Ed.), *Creativity and Collaborative Learning: A Practical Guide to Empowering Students and Teachers* (pp. 31-43). Baltimore, MD: Paul H. Brookes.

Krashen, S. D. (1982). *Principles and Practices in Second Language Acquisition*. Oxford, England: Pergamon Press.

Krashen, S. D. (1985). *The Input Hypothesis: Issues and Implications*. London/New York: Longman.

Kukulska-Hulme, A., & Shield, L. (2007). *An overview of mobile assisted language learning: can mobile devices support collaborative practice in speaking and listening?* Paper presented at the EuroCALL 2007, Coleraine, UK.

Liang, Y. (2004). The characteristics of Chinese character components and Chinese as a Foreign Language teaching. *Language Education and Research, 6*, 76-80.

Looi, C.-K., Wong, L.-H., So, H.-J., Seow, P., Toh, Y., Chen, W., et al. (2009). Anatomy of a mobilized lesson: Learning my way. *Computers & Education, 53*(4), 1120-1132.

Looi, C.-K., Wong, L.-H., & Song, Y. (in-press). Discovering Mobile Computer Supported Collaborative Learning. In C. Hmelo-Silver, A. O'Donnell, C. Chan & C. Chinn (Eds.), *The International Handbook of Collaborative Learning*. New York: Routledge.

Nussbaum, M., Alvarez, C., McFarlane, A., Gomez, F., Claro, S., & Radovic, D. (2009). Technology as small group face-to-face Collaborative Scaffolding. *Computers & Education, 52*(1), 147-153.

Shen, H. H. (2002). Level of cognitive processing: Effects on character learning among non-native learners of Chinese as a foreign language. *Language and Education, 18*, 167-182.

Tan, L., & Peng, D. (1991). The visual recognition process of the Chinese characters. *Psychology, 23*(3), 272-278.

VanPatten, B. (2004). Input and output in establishing form-meaning connections. In B. VanPatten, J. Williams, S. Rott & M. Overstreet (Eds.), *Form-meaning connection in second language acquisition* (pp. 29-47). Mahwah, NJ: Erlbaum.

Wang, P., Perfetti, C. A., & Liu, Y. (2003). Alphabetical readers quickly acquire orthographic structure in learning to read Chinese. *Scientific Studies of Reading, 7*(2), 183-208.

Wong, L.-H., Chai, C.-S., & Gao, P. (2010). Singapore primary school students' perceptions and challenges faced in using Chinese computer input systems for compositions. *Global Chinese Journal on Computers in Education, 6*(1-2), 46-61.

Wong, L.-H., Chin, C.-K., Tan, C.-L., & Liu, M. (2010). Students' personal and social meaning making in a Chinese idiom mobile learning environment. *Educational Technology & Society, 13*(4), 15-26.

Wong, L.-H., & Looi, C.-K. (2010). Vocabulary learning by mobile-assisted authentic content creation and social meaning-making: Two case studies. *Computer Assisted Learning, 26*(5), 421-433.

Zhu, Y., & Hong, W. (2005). Effects of digital voiced pronunciation and stroke sequence animation on character memorization of EFL learners. *Chinese Language Teachers Association, 40*(3), 49-70.

Zhu, Z. (2004). Chinese character structures and Chinese as foreign language teaching and learning. *Language Teaching and Research, 4*, 35-41.

Zurita, G., & Nussbaum, M. (2004). Computer supported collaborative learning using wirelessly interconnected handheld computers. *Computers & Education, 42*(3), 289-314.

NumberNet: Using Multi-Touch Technology to Support Within and Between-Group Mathematics Learning

Andrew Hatch, Steve Higgins, Andrew Joyce-Gibbons, Emma Mercier, Durham University Technology Enhanced Learning Lab, Leazes Road, Durham, DH1 1TA, UK.
Email: andrew.hatch; s.e.higgins; andrew.joyce-gibbons; emma.mercier@durham.ac.uk

Abstract: In this paper we present a new tool, NumberNet, designed to promote within and between group collaboration in a mathematics classroom. The activity builds on a standard individual mathematics activity (Explode-A-Number; Atkinson, 1992), to create a three-stage collaborative activity that promotes flexibility with numbers, operators and calculations. This tool uses a network of multi-touch tabletop computers in a classroom environment, taking students through a small group production activity, rotation of the activity between groups, and finally a sorting and structuring activity. Pilot results from 32 students indicate significant gains in the number of calculations that students produce from pre to post test. Further work and implications are described.

Introduction

Learning to use numbers in a flexible manner, and recognizing the breadth of operators that can be used and the different ways calculations can be constructed, is a key feature in mathematics education at the upper primary level in the United Kingdom. These skills establish a basis that prepares students for the more complex mathematical constructs and concepts that they will encounter during high school, such as algebra and geometry. However, while it is relatively easy to assess flexibility, designing curricula and activities to teach and foster numerical flexibility and adaptive expertise (Baroody & Dowker, 2003) is more challenging. In this paper, we describe a tool, NumberNet, that uses computer-supported collaborative learning activities to foster mathematical flexibility and reasoning through an interlinked series of small group and whole class activities. This tool builds on the standard classroom activity 'Explode-A-Number' (Atkinson, 1992), in which students are given a number and told to create as many calculations as possible that form that number (See figure 2 for an example of students work on a pre-test version of this task).

Although adult models of arithmetic calculation are relatively well known, the development of children's calculation skills is less well understood (for a recent overview see Kaufmann & Nuerk, 2005), and although the importance of developing flexibility through derived facts and mathematical reasoning is acknowledged, the relationships between mathematical concepts and processes is complex (Gray, 1994). It is generally agreed that derived fact and reasoning strategies are indicative of more successful mathematical learners (Baroody & Dowker, 2003) and, although teaching activities to develop these in computer-supported environments have not always been successful (Higgins, 2003; Sarama and Clemens, 2009) and the importance of experiences of teacher-supported mathematical activity in lessons is clearly essential (Ruthven, 1998)

Collaborative learning and problem solving activities have long been studied and recommended for teaching mathematics at many levels (see Esmonde, 2009; O'Donnell, 2006 for reviews). Research on collaborative learning indicates that the task-type and structure (e.g. Johnson & Johnson, 1991; Slavin, 1986), the quality of interaction, help seeking and problem-space creation can influence learning outcomes (Barron, 2003; Webb & Farivar, 1994; Roschelle, 1992). Bringing this research together, indicates the value of developing tools to support collaborative mathematics activities in the classroom, while paying close attention to how those activities are implemented and supported within the classroom. Adapting multi-touch technology for classroom use allows for the increased use of collaborative learning in the classroom.

Multi-touch technology provides a surface that can detect multiple touches, from one or more people, which creates opportunities for interaction between learners with multiple points of control. This is in contrast to single point of control computer technology that is more commonly used, and requires negotiation over control of the device (e.g. desktop or laptop computers that use a mouse or keyboard). In this study, we use a multi-touch enabled classroom (see Figure 1), which consists of four networked multi-touch student tables, an orchestration desk and a multi-touch interactive whiteboard. The student tables are designed to be sit-to-use, with space for between three and six students per table. The tables use a vision system that identifies up to thirty simultaneous touches, allowing all students to interact with the content through touch directly. The tables are networked, allowing for the teacher or student to 'slide' or send content from the orchestration desk to the student tables, or between the student tables. Finally, the orchestration desk can be used to move content from the student tables to the multi-touch interactive white-board. This means that teachers can project the content from a table to the whole class, using it to prompt a whole-class discussion about the methods or organization being used by an individual group. The teacher or students can interact directly with the projected content, explaining reasoning or pointing out difficulties to the whole class.

Previous work on collaboration with multi-touch tables suggests that when multi-touch is compared to single-touch, the use of multi-touch results in more task focused conversation (Harris et al., 2009), more equitable participation of collaborators (Marshall et al., 2009) and more speedy conflict resolution (Hornecker, Marshall, Dalton, & Rogers, 2008). When compared to paper-based tasks, the use of multi-touch tables also appears to promote joint attention and the creation of a joint problem space (Higgins et al, under review). Levering this technology for whole-classroom learning provides the opportunity for increasing support of small-group learning, and better ways for the teacher to orchestrate the tasks and moves between individual, small group and whole class discussion (Dillenbourg & Jermann, 2010).

NumberNet was designed to use the affordances of multi-touch to help students become more flexible in their use of mathematics. The tool builds on the multi-touch features, to allow small groups of students to work towards a joint goal of creating calculations for a single target number, while retaining control of their own input devices, similar to numeric keypads. The tool also takes advantage of the networking aspects of this technology, allowing the teacher to move the target numbers between tables. In this way, students can learn from their immediate group, from the calculations passed to them from other groups, and through whole class discussion towards the end of the activity.

In this paper, we present the tool and data from a pilot study, to explore the value of this tool for increased flexibility generating arithmetic calculations in primary school children.

Figure 1. The Multi-touch Enabled Classroom.

Method

As this was a pilot study of the use of the new NumberNet tool, the study was designed to collect both pre and post data of math skills, and as a participant observation study to explore how students used the tool and any issues that arose in relation to the collaboration, math or technology.

Participants were 32, year six pupils (10-11 year olds; mean age 10.5 years; SD = 0.35), who attended two local public primary schools during the 2010-11 academic year. Eight male and eight female students were recruited from each school; in both schools, the students recruited were all part of one class group, and so were familiar with each other and used to working together.

Students were recruited during a visit to the school by members of the research team about a week before they visited the lab. The purpose of this visit was to explain the study and the multi-touch technology and lead the students through a number of activities to prepare them for the visit to the lab; the students also completed the pre-test (described below) during this visit to the school. All students in the class (approximately 25) participated in the introductory visit. Parental consent forms were distributed and the teacher selected 16 students to visit the lab from those who returned consent forms.

Students were brought to the lab in a group of 16, spending about 5 hours working on a variety of learning activities and games on the tables. These activities included a number of non-educational game-type activities to familiarize the students with the technology, two history activities and four mathematics mysteries. All of the math tasks were word-based problems; one was a logic problem, another required sorting number facts to find an answer, and two required calculations. These tasks were based of classroom based tasks, and were unlikely to influence the students' performance on the post-tests. The NumberNet activity was introduced after lunch, at which stage the students had been working on the tables for a number of hours, and were familiar with interaction and operation of the tables. After using NumberNet, the students piloted other new tasks for the non-educational aspects of this project, including a sorting task where they had to pass photographs between the tables, and a three-dimensional version of Tetris; after approximately an hour, the post-test was distributed.

The data described in this paper consists of comparison of the pre and post-test data, and an exploration of the process of one group. As this was a pilot study, each of the authors of this paper were present during this activity, acting as participant observers to evaluate the task; the groups at each table were each recorded by two cameras and a recording of the computer display was made. Participant observers were focused on helping the children complete the task while paying attention to any difficulty that they had with the task and the strategies that the group members used. During a de-briefing session after each day of data collection, the participant observers recounted what they had noticed, paying particular attention to processing of knowledge exchange, helping behaviour or on-task conversations within their group. The focus group was selected through the comments of the participant observers, where the sharing of a strategy had been noticed; the video of this group was reviewed in preparing the analysis for this paper. Further data from a larger study will be presented during the conference.

Pre and Post Test

Students were given a pre-test during the initial classroom visit, and a post-test about an hour after they used the NumberNet tool in the lab. The test consisted of a single sheet of paper, with space for the student's name, and the statement: My target number is x, with x being replaced with one of up to ten possible numbers (six at pre-test and four at post-test). Different target numbers were used for the pre and post-tests, and no two children at the same table received the same target numbers.

The tests and pencils were distributed to each student, and they were asked to put their names on the top of the page. There were then given brief instructions (all students were familiar with the task from prior use in their mathematics class), and told they had two minutes to write as many calculations as possible. After two minutes, the students were asked to put down their pencils and their tests were collected by members of the research team.

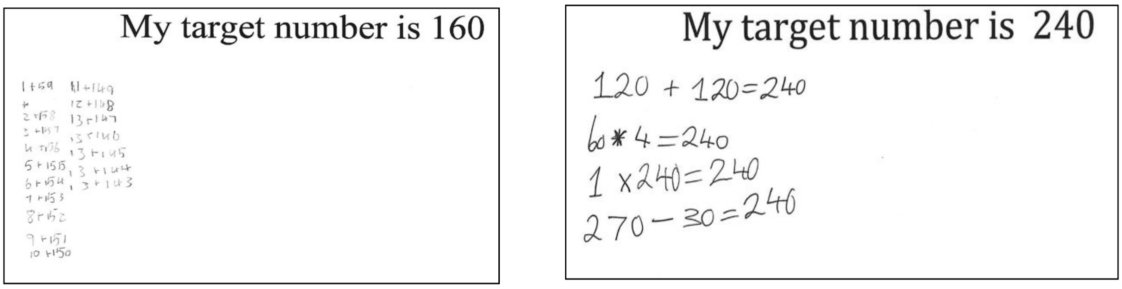


Figure 2. Examples of the Pre and Post-test (Standard Explode-A-Number Activity).

NumberNet

The NumberNet activity is based on a standard classroom activity known as 'Explode a Number' (e.g. Atkinson, 1992). This task is commonly used in upper primary school (9-11 year olds) in the mental/oral arithmetic section of the mathematics lesson. Students are given a number and told to create as many alternative calculations as possible, which give that total, usually within a certain time limit. The task can be varied depending on the students' ability, areas in which they need practice and the learning objectives for the activity by specifying which operations can be used to achieve the answer. Over the course of time, students are expected to identify a variety of strategies for this task, including using simple addition and subtraction patterns (e.g. 499+1=500; 498+2=500; 497+3=500 etc), simple multiplication or division patterns (e.g. 50*10=500; 5*100=500; 0.5*1000=500; 5000/10=500 etc), multiplication or division by one (e.g. 500/1 = 500; 500*1= 500), commutative statements (e.g. 200+300=500; 300+200=500; 250*2 =500; 2*250=500). The overall goal is for students to be able to identify the different uses of operations and develop a flexible understanding of how numbers and operators interact.

While this task provides an easy way of assessing a child's current understanding of numbers and operators, and can be used over time as a practice activity, it is explicitly an individual activity. While some teachers attempt various whole-class extension activities after students have written their calculations, it remains difficult to adapt this from a practice or assessment activity into one that provides opportunities to learn to develop a wider range of strategies.

The NumberNet tool takes the principle of the task, but by making it a collaborative task, and allowing for an extension sorting activity, provides a way for students to learn from each other during production of calculations, and for the teacher to structure and scaffold their learning in the conversations that follow. The three phases of NumberNet take the students through small group, between group and whole class interactions, allowing them to learn from their group members as well as other teams and from the classroom discourse.

Phase 1

In the first phase of NumberNet a target number is assigned to each table (a different number for each table) and each student at the table has access to a multi-touch numeric keypad on the surface of the table (see Figure 2) that is designated for their use. The teacher can choose when to make the target numbers appear, and how long the students are allowed to spend creating calculations for that number. Each student can work alone, or can discuss with their peers at the table how to address this task or particular calculations within the task.

The numeric keypads allow the students to use the digits 0-9, the four primary operators (addition, subtraction, multiplication and division) and brackets (parentheses). While the teacher controls allow for the removal of any digit or operator on the numeric keypads (to make the task more difficult or encourage certain students to focus on a particular operator, factors etc), these features were not used during the experiment described here: all students were able to use all digits and operators. When the student has created a calculation, they use the send button to move the calculation from the number pad and onto the open space on the table.

The keypads are set to identify duplicate calculations (e.g. calculations that are already on the table, regardless of who created them), turning them red and not allowing the student to slide it into the middle of the table. Incorrect calculations and invalid expressions are permitted at this stage, although are flagged in the teacher's control panel.

Once a calculation is added to the table, any student at that table can pull it onto their number pad and edit it, allowing for corrections or discussion within the group. The teacher's control records which number pads create each calculation, and whether the calculations are new or edits to existing calculations. This provides a way for the teacher to monitor the task from a distance during the task, or, what is more likely, examine the activity afterwards, or over time. This information can be used to identify students who appear to be stuck on a particular method, students who alter calculations during the tasks, and students who do not get the calculations correct. Over time, this can provide the teacher with a method of assessing the students' performance, learning and areas of difficulty and indicate where prompting exploration of additional strategies would be useful.

Throughout the task, the total number of correct calculations on each table is displayed in the corner of the screen. This feature can be turned on or off at any time during the activity, depending on whether the teacher is interested in encouraging between-group competition. For the pilot study described here, this feature was enabled, allowing the students to see which table was creating the most calculations.

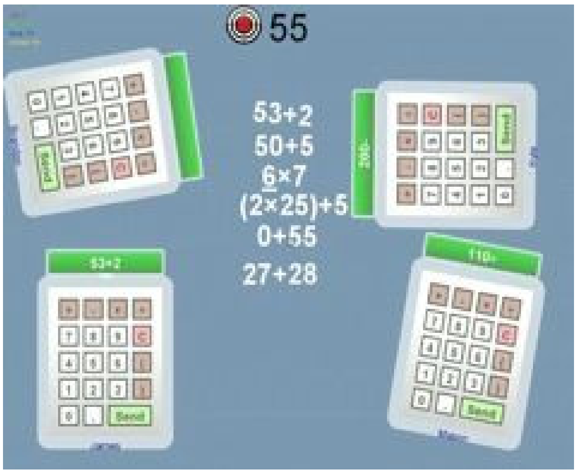


Figure 3. Phase One of NumberNet, during Creation of Calculations.

Phase 2

After a length of time determined by the teacher (either before or during the task), the second phase of NumberNet begins. The number pads disappear, the correct calculations cluster near the top of the table, while the incorrect calculations remain in the center of the table; the teacher can allow the group to examine their incorrect calculations before moving onto the next part of the task. The correct calculations and the target number are then rotated to the next table, so that each group receives the last group's number and their calculations. They are given a few minutes to explore the calculations that have been sent to their table, and then the number pads appear so they can add more calculations. The length of time before the number pads appear can be altered, and increases for each rotation, as the students have increasing numbers of calculations to explore. As before, the tool identifies duplications, so the new group cannot replicate calculations created by the previous group.

Phase 3

Once the target numbers have been rotated around all the tables (four tables in our case), the teacher can activate the final phase of the activity. In this phase, the groups get their original target numbers back, with all the calculations that have been created by their class. The number pads disappear and the group's task is to link the calculations together in a way that makes sense to them (see Figure 3). This can then be projected by the teacher, so that the whole class can discuss the reasons for linking and what the similarities or differences are between calculations. There are many possible ways to create structures for the calculations (e.g. similar patterns, the same operators, etc), the important feature is that the students can explain why they linked particular calculations, and, over time, understand the range of possibilities.

Calculations are linked together by placing a finger on each one and bringing them together; when the two calculations touch, a link is formed. Links can be broken by pulling the calculations apart. Groups can enable or disable a dynamic automatic layout of the resulting graphs of expressions. The layout algorithm used is a simple force-directed graph layout that attempts to minimize edge crossings.

Due to technical issues, not every group in our sample was able to use Phase Three, and the teacher noted that additional work will be necessary to develop pedagogical supports for the classroom discussion of this activity; however changes to the technology have resulted in increased stability for this phase, and the data presented at the conference will include use of Phase Three.

Technical Details

Each student table ran an instance of the NumberNet student software which was built on our software framework which was written to support distributed applications for the classroom. This framework uses Java with OpenGL bindings through the jME2 scene graph library. Our framework specifically provides components for touch interaction, content management and teacher control over the network. Inter-table messaging is achieved with opens standards messaging systems (XMPP). The teacher's software for NumberNet was built using Multiplicity with a Swing GUI.

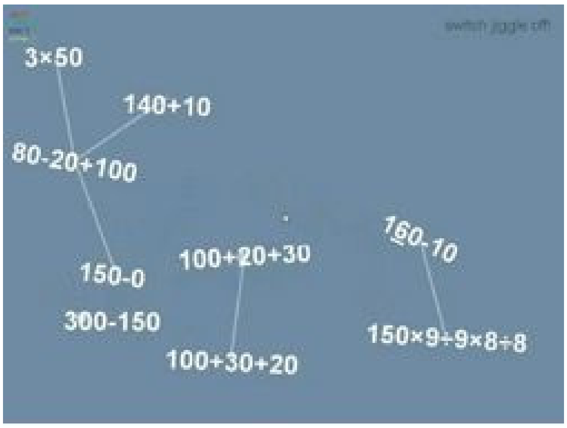


Figure 3. Phase Three of NumberNet: Sorting the Calculations.

Results

Outcome Measures

The data from the pre and post-test were collected and analyzed to explore whether there were differences in number of calculations created, range of operators used and maximum number of operators use in a single calculation. Paired sample t-tests were used to determine whether these scores differed significantly from pre to post test.

Results indicated that the difference in number of calculations created from pre to post test was statistically significant, $t(31) = -3.63$, $p < .001$. The differences in range of operators was not significant, $t(31) = .55$, $p=.59$. The difference in maximum number of operators used was not significant, $t(31)=-1.53$, $p = .14$. Descriptive statistics are displayed in table 1.

Table 1: Descriptive statistics for number of calculations, range of operators and maximum number of operators.

	Pre-test		Post-test	
	Mean	SD	Mean	SD
Number of Calculations	5.91	3.49	8.59	4.63
Range of Operators	2.5	1.14	2.41	.84
Max number of operators	1.38	.94	1.56	1.13

Within-group Learning

While our data indicated an increased number of calculations created, but not an increase in the maximum number of operators used, or an increase in the range of operators used, through exploration of one group's process, we can see some changes to their calculations during the task, shedding light on the mechanisms through which NumberNet might support the developmental of mathematical understanding.

The focus group is made up of two boys, Jack and Thomas, and two girls, Olivia and Amy. This group represent a wide range of mathematical skills, ranging from Jack, who produced two correct (and one incorrect) calculation at pre-test to Thomas who produced seven correct calculations at pre-test. The group also included Olivia, who only produced three calculations at pre-test, although one of these was a 4-operator calculation, one of the longest calculations seen in the entire data-set and the only multi-operator produced at pre-test by the members of this group. At pre-test, Amy produced five correct and one incorrect calculation. The calculations created at pre and post tests (as well as the target numbers) are shown in table 2.

There were two primary changes to the group's post-test data that can be tracked back to the group's process. The first is the introduction of multi-operator expressions. During the course of the activity, Olivia created a number of multi-operator calculations; members of her group were then seen to also create multi-operator expression, possibly as a reaction to seeing her use this form of calculation (usually just two operators, without accurate use of brackets/parentheses). At the post-test, we see Amy produce a multi-operator calculation (6*8*10), suggesting the possibility that she had picked this up from observing Olivia's behavior during the activity.

The second change that we saw was in Jack's calculations from pre to post test. At pre-test, Jack created two correct and one incorrect calculation. His pre-test calculations contained one correct subtraction and one correct addition, at post-test this increased to three correct additions, two correct subtractions and a multiplication calculation (all single operator expressions). Three of these are also small manipulations around his target number (e.g. 499+1=500; 500-0=500). As this type of pattern was created by other members of his group, it suggests that he may have noticed that he could create these simple calculations, and used this strategy as a way to create more calculations at post-test.

Table 2: Calculations created by each of the focus group students at pre and post (not all are correct).

Jack		Thomas		Olivia		Amy	
Pre (320)	Post (500)	Pre (320)	Post (160)	Pre (240)	Post (235)	Pre (180)	Post (480)
311+9	1+499	3200/10	1600-100	200+40	200+35	60*3	470+10
10-330	503-3	200+120	80*2	100+140	200+30+5	1800/10	479+1
20+300	10+490	20+300	16*10	50+50+50 +50+40	100+100+35	179+1	6*8*10
	300+200	310+10	100+60		50+50+50+50 +30+5	380-200	400+80
	100*5	1+319	17*10-10		100+100+20+ 10+4+11	174+6	500-20
	500-0	320+0	40+120		200+19+11+3 +2		438-3
		2+318	1+159				
			20+140				
			10+150				

Discussion

The results of our pilot study exploring the value of NumberNet for increasing mathematical flexibility and developing adaptive expertise in upper primary students indicated that there was a significant increase in the number of calculations created by students from pre to post test. This result suggests that from a single use of

NumberNet, improvements in flexibility with mathematical concepts were evident. Exploration of the process also indicated that these improvements may have been due to the collaborative nature of the activity.

While the results from the pilot study presented in this paper indicate the possibilities of NumberNet, other explanations include the chances that these occurred due to the opportunity of practicing creating calculations, rather than any direct effect of the tool. Further studies comparing students who used NumberNet with students who practiced the paper-based version of the activity are necessary to determine whether this was in fact the case. Further studies will also explore the role of the third phase of NumberNet in more detail to understand how the process of sorting or creating structure of calculations may foster numerical flexibility.

Additionally, the data presented in this study came from two pilot tests of the tool; each test was orchestrated by a different teacher. While both teachers have extensive in-classroom experience using the Explode-A-Number activity in a standard classroom and experience using the multi-touch enabled classroom; the NumberNet activity was new to them. They both expressed the need to develop the pedagogical features of this activity in more detail, helping the teacher to understand how best to manage and orchestrate the activity and scaffold students' conversations in the final stages of the task to take the most advantage of the tool. Future development will include the pedagogical supporting documents and further testing with a range of teachers. Finally, due to technical issues, the third phase of the activity could not be completed by every group; we anticipate that this phase will lead to more complex conversation around the activity, and that greater changes in flexibility and the range of calculations would be seen in groups who had the opportunity to fully engage with this part of the tool.

NumberNet was designed to take advantage of the affordances of a multi-touch classroom. in this way, the tool is an example of CSCL embedded in the classroom, which supports small group, intra-group and whole class learning and facilitates the movement between these phases for the teacher. It also provides evidence for the way in which integrating technology into the classroom can be used to introduce new types of interactions between students, and engagement with content and ideas. By leveraging the affordances of the technology to promote sharing of strategies in tasks similar to Explode-A-Number, students have the opportunity to learn from each other and recognize different ways of approaching a similar task.

Learning to be flexible users of knowledge has been described as adaptive expertise (e.g. Hatano & Inagaki, 1986), and recognized as an important skill to develop for future learning and transfer. To develop adaptive expertise, Schwartz, Bransford & Sears (2005), suggest that patterns of learning should balance the development of efficiency using strategies, with experiences of innovation, keeping learning with the 'Optimal Adaptability Corridor'. Evidence from prior research indicates that when students learn mathematics in ways focus on the development of efficiency, deep understanding of the constructs can be limited, while students who learn using a more innovative approach took longer to master the fundamental ideas but understood the concepts and were better prepared to transfer their knowledge into a new situation (Martin & Schwartz, 2005). In NumberNet, students can move between using their pre-existing efficient strategies for creating calculations and innovating new strategies through observation or collaboration with their peers during the initial stages of the task. The third phase requires all students to innovate to structure the calculations that were created. The teacher can guide this activity in a number of ways, and provide support for the whole class through projection of each group's calculation maps. In future work, the use of NumberNet over time will be explored, to determine whether it promotes the development of flexibility and adaptive expertise in primary mathematics.

References

Atkinson, S. (1992) Mathematics with Reason: The Emergent Approach to Primary Maths London: Heinemann.

Baroody, A.J. & Dowker, A. (2003) The development of arithmetic concepts and skills: constructing adaptive expertise London: Routledge

Barron, B. (2000). Achieving Coordination in Collaborative Problem-Solving Groups. Journal of the Learning Sciences, 9(4), 403-436. doi: 10.1207/S15327809JLS0904_2.

Barron, B. (2003). When Smart Groups Fail. Journal of the Learning Sciences, 12(3), 307-359. doi: 10.1207/S15327809JLS1203_1.

Dillenbourg, P., & Jermann, P. (2010). Technology for classroom orchestration. In M. S. Khine & I. M. Saleh (Eds.), *New Science of Learning* (pp. 525-552). New York, NY: Springer New York.

Esmonde, I. (2009). Ideas and Identities: Supporting Equity in Cooperative Mathematics Learning. Review of Educational Research, 79(2), 1008-1043. doi: 10.3102/0034654309332562.

Hatano, G., & Inagaki, K. (1986). Two courses of expertise. In H. Stevenson, H. Azuma, & K. Hakuta (Eds.), Child development and education in Japan (pp.262–272). New York: Freeman.

Harris, A., Rick, J., Bonnett, V., Yuill, N., Fleck, R., Marshall, P., et al. (2009). Around the table: are multiple-touch surfaces better than single-touch for children's collaborative interactions?. Proceedings of the 9th international conference on Computer Supported Collaborative Learning-Volume 1 (p. 335–344). International Society of the Learning Sciences. Retrieved from http://portal.acm.org/ citation.cfm?id=1600053.1600104.\

Higgins, S., Burd, E., Mercier, E. & Joyce-Gibbons, A. (under review) Multi-touch tables, pedagogic design and knowledge transformation. Journal of Computer Assisted Learning.

Higgins, S. (2003) Does ICT make mathematics teaching more effective? in I. Thompson (Ed) Enhancing Primary Mathematics Teaching Buckingham: Open University Press

Hornecker, E., Marshall, P., Dalton, N., & Rogers, Y., (2008). Collaboration and Interference: Awareness with Mice or Touch Input. Proceedings of CSCL, 167-176.

Johnson, D. W. & Johnson, R. T. (1991) Learning together and alone: Cooperative, competitive and individualistic learning. Englewood Cliffs, NJ. Prentice Hall

Martin, T., & Schwartz, D. L. (2005). Physically Distributed Learning: Adapting and Reinterpreting Physical Environments in the Development of Fraction Concepts. *Cognitive Science*, *29*(4), 587-625. doi: 10.1207/s15516709cog0000_15.

Marshall, P., Fleck, R., Harris, A., Rick, J., Hornecker, E., Rogers, Y., et al. (2009). Fighting for control: Children's embodied interactions when using physical and digital representations. Proceedings of the 27th international conference on Human factors in computing systems (p. 2149–2152). ACM New York, NY, USA.

O'Donnell, A. (2006). The role of peers and group learning. In P. Alexander & P. Winne (Eds.), The Role of Peers and Group Learning (2nd ed., pp. 781-802). Mahwah, NJ: Lawrence Earlbaum.

Roschelle, J. (1992). Learning by collaborating: convergent conceptual change. The Journal of the Learning Sciences, 2(3), 235-276.

Ruthven, K. (1998) The Use of Mental, Written and Calculator Strategies of Numerical Computation by Upper Primary Pupils within a 'Calculator-aware' Number Curriculum British Educational Research Journal, 24.1: 21-42.

Sarama, J. and Clements, D.H. (2009) Early childhood mathematics education research: learning trajectories for young children London: Taylor & Francis

Schwartz, D. L., Bransford, J. D., & Sears, D. (2005). Efficiency and innovation in transfer. *Transfer of learning from a modern multidisciplinary perspective* (p. 1–51).

Slavin, R.E. (1986) Using student team learning. (3rd ed.) Baltimore, MD: John Hopkins University.

Webb, N. M., & Farivar, S. (1994). Promoting helping behavior in cooperative small groups in middle school mathematics. American Educational Research Journal, 31(2), 369.

Collaborative Gaze Footprints: Correlates of Interaction Quality

Patrick Jermann, CRAFT, Ecole Polytechnique Fédérale de Lausanne (EPFL), Patrick.Jermann@epfl.ch
Dejana Mullins, University of Freiburg, D-79085 Freiburg, dejana.mullins@gmail.com
Marc-Antoine Nüssli, Pierre Dillenbourg, CRAFT, Ecole Polytechnique Fédérale de Lausanne (EPFL),
Email: Marc-Antoine.Nuessli@epfl.ch, Pierre.Dillenbourg@epfl.ch

Abstract: Dual eye tracking offers new possibilities for the analysis and diagnosis of collaborative interaction. Cross-recurrence analyses and visualizations offer insight into how closely two collaborators' gaze follow each other. We contrast two cases to illustrate how gaze cross-recurrence can be used as a correlate of high and low quality interaction. The intriguing graphical patterns that result from the time coupled traces of the collaborators' fixations are footprints of the quality of the interaction. Good quality interaction features a higher recurrence rate than low quality interaction. The graphical structure of the recurrence plots indicates whether collaborators divide labor and whether they are sharing visual attention.

Introduction

The long-term goal we pursue is to use behavioral indicators that can be measured and evaluated automatically to represent the quality of interaction along psychologically meaningful dimensions. These dimensions could then be used as a basis to diagnose the quality of interaction (Soller, Martinez, Jermann & Mühlenbrock, 2005).

We contribute to this endeavor by pointing out gaze-based correlates of rating dimensions that are used in CSCL to characterize the quality of collaborative interaction (Meier, Spada, & Rummel, 2007). The assumption behind this approach is that the type of coupling between collaborator's gaze reflects the quality of their interaction. We focus in this contribution on cross-recurrence, a measure of whether two collaborators look at the same target more or less at the same time (typically within a 4 seconds time span). We see cross-recurrence visualizations (Figure 1) as footprints of the quality of interaction that allow to visually grasp the gist of the interaction dynamics. More generally, our approach aims at describing and modeling collaboration across several levels of control (Lord & Levy, 1994), from low-level individual signals up to social interaction via an analysis of synchronization and alignment between individual behaviors.

Dual Eye Tracking and Gaze Recurrence

Synchronous dual eye tracking is a novel methodology, which consists of recording the gaze of two collaborators simultaneously (see for instance Cherubini, Nüssli & Dillenbourg, 2008 and Nüssli, Jermann, & Dillenbourg, 2009). Existing research about gaze and communication studied situations that are often about "delayed" communication (subjects describe a scene for some listener who is not present) and about simple visual referents (the scene looked at by the subjects is simple and static). Richardson and Dale (2005) for example used eye movements to investigate how a speaker and a listener deployed their attention within a visual "common ground". Their research focused on cases where two partners are looking at a visual scene that is the topic of conversation. A cross-recurrence analysis showed that speaker and listener eye movements were coupled throughout the discourse. The more closely a listener's eye movements were coupled with the speaker's, the better the listener did on comprehension questions. In this study, speaker and listener were not interacting synchronously. The result was however replicated in a study with synchronous dialogue (Richardson, Dale, & Kirkham, 2007). Cross recurrence quantification analysis was proposed recently as a generalized method to unveil the interlocking of two interacting people (Dale, Warlaumont & Richardson, in press). The authors used the method to analyze gaze as well as the use of language alignment through grammatical sequences (Dale & Spivey, 2006). In previous studies conducted in our lab, we found that gaze proximity was, at the micro-level, negatively related to the misunderstandings of referential utterances (Cherubini et al., 2008) and, that cross-recurrence was at the macro-level, positively correlated to team performance (Cherubini, Nüssli & Dillenbourg, 2010). Some initial observations by Pietinen and colleagues (2008) in the domain of pair programming also suggest that gaze closeness might reflect tightness of collaboration. The driver and navigator roles that are typical for pair programming seem also to be reflected in the way programmers look at the code.

We extend the work by these authors on pair programming, a task that is more complex than controlled referential communication and more realistic than faked interaction. The notion of footprint that we use in this contribution is related to the notion of "signature" that is used by Dale et al. (submitted) to refer to the high level visual patterns (lines and structures) in cross-recurrence plots. In physics, where the cross-recurrence plots initially stem from, Eckman Kamphorst, and Ruelle (1987) also describe two levels of reading recurrence plots: "large scale typology" which refers to the overall aspect of the plots as containing structure or being homogeneous; and "small scale texture" which refers to smaller details and reflects intrinsic properties of the dynamical systems represented.

Collaboration Quality

CSCL encompasses diverse approaches for analyzing interaction that differ in the purpose of the analysis, the units of interaction that are analyzed, and the underlying theoretical assumptions. In a recent workshop series (e.g. Suthers, Law, Lund, Rose, & Teplovs, 2009), this diversity and the importance of achieving a common ground in the analysis of interaction was specifically addressed.

Traditionally, researchers applied a "coding and counting" approach in order to evaluate the effectiveness of collaboration processes (e.g. De Wever, Schellens, Valcke, & Van Keer, 2006; Strijbos, Marten, Prins, & Jochems, 2006), that is, they transcribed the interaction and counted the occurence of particular types of utterances. This approach is not only very time consuming, but the results of a coding analysis also do not necessarily inform about the success of the collaboration. For instance, a high amount of utterances concerning the technical coordination can also indicate that the collaboration partners have difficulties in coordinating the use of the technical environment. Meier et al. (2007) therefore suggested employing a rating analysis. They developed a rating scheme to analyze the quality of collaborative interaction. The rating scheme contained nine dimensions that evaluated aspects of communication, joint information processing, coordination, interpersonal relationship, and motivation. The evaluation of the rating scheme testified a high inter-rater reliability, consistency, and validity. Originally, the rating scheme was developed for the context of interdisciplinary collaboration. However, Rummel, Deiglmayr, Spada, Kahrimanis, and Avouris (2011) showed that it can also be adapted for other contexts, for instance for the analysis of collaborative interaction between computer scientists, or for the analysis of students' collaborative learning in mathematics (see also Diziol et al., 2008).

Research Question

The goal of this contribution is to point out gaze-based correlates of the dimensions of collaborative interaction identified by Meier et al. (2007). To this end, we present a case study that aims at identifying the characteristics of gaze traces that are related with interaction quality. In order to identify relevant gaze features, we contrasted two cases: One dyad that showed high collaboration quality and one dyad that showed low collaboration quality based on the interaction analysis with a rating scheme. By choosing diverse examples, we exploit the natural variability between the cases to identify relevant gaze patterns (Firestone, 1993). Our main question is what are the differences in the gaze patterns of the good and the bad dyad, and how do these differences relate to the interaction quality?

Method

We report data that was collected during the piloting phase of a larger study of program understanding that involved 42 pairs and 55 individuals, all students in computer science and communication science, and which is under analysis at the time of writing. We focus specifically in this contribution on three master students in computer science in their 4th and 5th year and one bachelor 3rd year student in communication science. They all report to program at least between 1 to 3 hours per week in Java. We refer to the participants as "Blue" and "Red". In Dyad 1, Blue indicated that he has "poor" knowledge of Java, whereas Red indicated that he has "good" knowledge. In Dyad 2, Blue indicated "good" knowledge and Red indicated "very good" knowledge. Despite this self-reported disparity of expertise, the dyad partners had comparable prior knowledge as assessed in a pre-test: In both dyads, one student reached a prior knowledge score of 5 out of 13 points (Blue in both dyads), and the other student reached a prior knowledge score of 7 points (Red in both dyads).

Dyads worked on a task that consisted of describing the rules of a game implemented as a Java program of three hundred lines. Understanding and explaining the rules of the game required subjects to translate programmatic Java code into a domain model that is expressed in everyday language. In the game, two players alternated in choosing numbers out of a list of nine numbers from 1 to 9. The goal of the game was to have three numbers that add up to 15. The instructions for the dyads were: "You have been hired in a game company and you have received this code. This is a two player turn-based arithmetic game. Please explain the rules to potential non-programmer players including: What is the initial game situation, what does a player have to do at each turn. Give an example of a valid and an invalid action. What are the general rules of validity of an action? What does a player have to do or get in order to win the game? Under which condition can the game end without a winner?"

The experiment consisted of two phases. First, students completed an individual pre-test to assess their prior knowledge in programming and more specifically their java skills (individual phase). Then they collaboratively solved ten programming tasks (collaborative phase). During the collaboration, the dyad partners were seated in separate rooms, but were able to communicate with each other through an audio channel. Each question was timed, and the dyads received a warning 20 to 45 seconds before time ran out. Although we recorded the entire interaction, the dyads were asked to indicate when they were ready to answer to the tasks by clicking a button labeled “Start recording”. The entire experiment lasted for approximately 90 minutes.

The problem-solving environment consisted of a collaborative programming editor based on the Eclipse IDE (http://www.eclipse.org). The workspace consisted of an instruction view (Instructions area) at the top of the screen that contained the question the subjects had to answer, and a central view (Code area) displaying the code of the program. The lower part of the editor displayed the remaining time and contained a button for recording the answer (Time/Management area). The workspace was shared via a server that relayed interface events to connected clients, i.e. if one partner used the scroll bar, the view of his partner changed as well to ensure that they were viewing the same section of the code. Furthermore, the dyad partners were able to highlight parts of the code in order to draw the partner’s attention to specific aspects or sections. Two Tobii 1750 eye-trackers were used to record gaze at 50 Hz. A dedicated server synchronized the eye-trackers’ recordings, and the data was logged using callback functions from low-level API of the eye-trackers. The subjects’ heads were held still with a chinrest placed at 65 centimeters of the screen. An adaptive algorithm was used to identify fixations and a post-calibration was done to best align fixations with the stimulus. While subjects used Skype to converse, audio was recorded using dedicated phantom powered boundary microphones amplified through a preamplifier.

Analysis of Collaboration Quality

To evaluate the collaboration quality of dyadic pair programming, we replayed the interaction of the dyads using an interactive tool, which displays the shared screen, the selections done by the subjects, and plays back their conversation and eye gaze. The analysis was two-folded: First, we rated the collaboration quality with a rating scheme that is based on the work of Meier et al. (2007; see also Rummel et al., 2011). Second, we compared the dyads’ eye gaze during various moments of the interaction to identify whether gaze indicators are sensitive to contrasts in the rating dimensions.

Rating Analysis

As the main goal of the current study was to identify simple gaze-based correlates of collaboration quality, we concentrated the analysis on two aspects of collaboration quality: communication and coordination (cf. Rummel et al., 2011). Two other aspects, joint information processing, as well as interpersonal relationship and motivation were not taken into account as these may have less clear indicators in dyadic gaze. We evaluated five dimensions that we briefly describe hereafter. Two dimensions assess the quality of the dyadic communication, and three dimensions assess the dyad’s coordination of joint work.

Collaboration flow evaluates if the collaboration partners engage in a coherent exchange of information and keep mutual awareness on what the partner is currently working on. High collaboration flow is marked by a seamless “flow” of dialogue while collaboratively solving the joint task, in other words, the dyad partners address each other and react to each other’s proposals or questions. The dimension *sustaining mutual understanding* evaluates if the dyad partners successfully work towards a shared basis of understanding, in other words, a common ground (cf. Clark & Brennan, 1991). Dyads that show successful grounding give and ask for feedback. Mutual understanding can also be facilitated if the dyad partners clarify about which aspect they will talk next, for instance by verbal references to the specific part of the code, or by selecting the specific part of the code in the shared workspace (deictic reference). The dimension *technical coordination* assesses if the dyad partners effectively use the technical environment during problem solving and capitalize on the system resources (cf. Diziol et al., 2008). In the current study, the dyad partners work on a joint work space: if one partner scrolls, the window of the other dyad partner changes as well; if one partner highlights code, his or her partner can see it. In contrast to Meier et al. (2007), the current study evaluates the coordination of the joint problem-solving with two separate dimensions. Similar to Meier et al. (2007), we evaluate the amount of *task division*, in other words: Do the dyad partners solve all aspects of the task together, or do they alternate between individual and collaborative work phases? In addition, the dimension *participation symmetry* assesses if both dyad partners are equally engaged in problem solving.

As in Meier et al. (2007), we assessed the dimensions on a 5-point rating scale from - 2 (low quality) to + 2 (high quality). The dimension task division was an exception. While phases of individual work are crucial for a successful outcome in *interdisciplinary* interaction (Rummel & Spada, 2005), it is not yet clear if a certain amount of task division is also beneficial in the current *pair-programming* setting, or if it is better if the dyad partners constantly interact with each other. We therefore evaluated the amount of task division with a rating scale from 0 (no task division) to 4 (high amount of task division). The two first authors of this paper separately rated the quality of the collaboration and reached a high agreement; disagreements in the ratings were solved by mutual discussion. Because of the small number of dyads, we cannot report kappa statistics about inter-rater agreement.

Cross-recurrence Plots

Cross-recurrence plots represent the time-dependence of two processes in a dynamical system. Because the graphical interpretation of cross-recurrence is not straightforward, we hereafter describe a schematic cross-recurrence plot along a real plot for illustration purpose (see Figure 1).

The horizontal axis represents time for the first collaborator and the vertical axis represents time for the second collaborator. Each pixel of the plot (Figure 1, right) corresponds to 200 milliseconds time slice (the duration of short gaze fixations are around 100 ms). For a pixel to be colored, the distance between the fixations of the two collaborators has to be lower than a given threshold (70 pixels in our case). Moments of recurrence appear colored in the plot (neither light gray nor white). Cross-recurrence plots show exclusively recurrent gaze. If two collaborators continuously looked at two different spots on the screen for the whole interaction, the resulting cross-recurrence plot would be completely blank (light gray in Figure 1, right). On the contrary, if the two collaborators looked at the same spot on the screen totally synchronously, the plot would show only a dark line on the diagonal. Pixels exactly on the diagonal of the plot correspond to synchronous recurrence, e.g. collaborators look at the "same" target at exactly the same time. Pixels below the diagonal show a lag for the first collaborator (the second collaborator looked at the target before) whereas the pixels above the diagonal show a lag for the second collaborator. Asymmetries above and below the diagonal line could therefore be indicative of leading and following behaviors.

In some cases, the colored areas on the real plot form rectangular shapes, which are represented by rectangles in the schematic plot (Figure 1, left). Labels refer to areas of interest on the screen (e.g. for a computer program *a* would typically refer to the method signature, *b* to the body of the method, and *c* to the return statement). On the horizontal axis, the labels of the arrows represent the areas of the stimulus that were looked at by the first collaborator. The corresponding sequence of areas of interest that were looked at is "*a->b->a->c->a*". The vertical axis represents the sequence of areas focused on by the second collaborator: "*a->c->b->a->c*". Following the diagonal, we move through the interaction, and we see that the dyad has looked together synchronously at the areas "*a->b->a->c*". Off-diagonal regions indicate asynchronous cross-recurrence, which happens when collaborators peek at or review places they have been or will be looking at together. For example, in the schematic plot (Figure 1, left), the area labeled *c'* indicates that the second collaborator has looked at *c* long before the dyad focuses on *c* together. The areas marked *a'* are indicative of the first collaborator looking back at area *a* after having looked at it together with the second collaborator on two occasions.

Hence, in the domain of program understanding, cross-recurrence plots not only represent the "togetherness" of collaborative code exploration, but also show whether the exploration of the program is rather linear (there is almost no off-diagonal activity because regions of code are looked at only once) or iterative (there are marked off diagonal regions that result from the collaborators repeatedly looking at the same regions of the code).

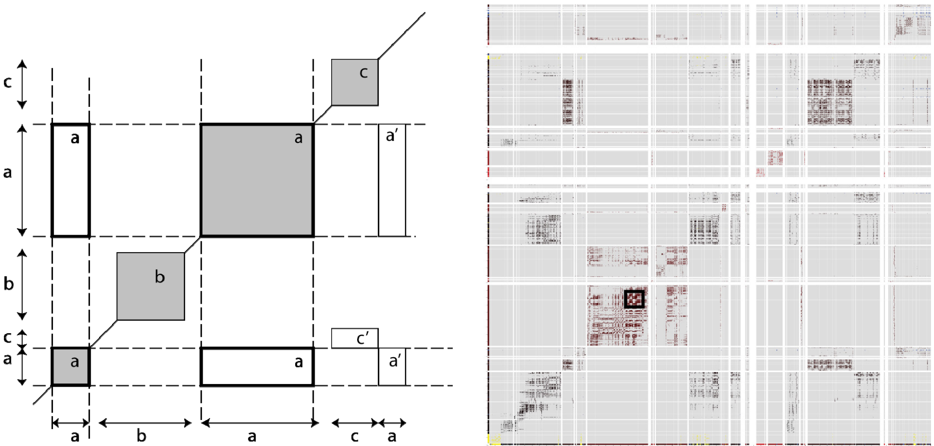


Figure 1. Schematic (left) and Real (right) Cross-recurrence Plots.

Results

Two dyads were chosen among available data at the time of writing because they best illustrated what is commonly understood in CSCL as a "good" and "bad" interaction quality. For each dimension of interaction quality (as defined by the rating scheme) we report a characterization of the interaction through gaze cross-recurrence plots and associated measures.

Gaze Correlates of the Dimensions of Interaction Quality

Overall, the collaboration quality of Dyad 1 is very bad. Their interaction is rather lopsided: Red explains to Blue how he understands the code, while Blue is hardly contributing to the discussion and back-channels only infrequently. For Red, it apparently was quite easy to understand the general functionality of the code, and except for a small mistake - he believes that the players can choose from a list of ten numbers instead of nine - he correctly explains the rules of the game. In contrast, Dyad 2 overall shows a high collaboration quality. After collaborators gained a good overview of the code, they jump to specific sections of the code in order to answer open questions. The dyad is able to explain the general functionality of the game.

In the remainder of this section, we describe the dyads' collaboration in more detail with reference to the dimensions of the rating scheme, and intertwine the descriptions with results from gaze analyses. Table 1 summarizes the ratings of the dyads' collaboration quality. Our goal in this contribution is to identify whether strong contrasts in ratings are reflected in behavioral measures.

Table 1: Collaboration quality of the two contrasting cases.

	Collaboration Flow	Mutual Understanding	Technical Coordination	Task Division	Particiation Symmetry
Dyad 1	-2	-2	-2	4	-2
Dyad 2	2	2	2	1	2

Collaboration Flow

The collaboration flow of **Dyad 1** is very bad. Mostly Red is talking, and he hardly reacts to Blue's attempts to interrupt his monologue. Only infrequently, Red addresses his partner to ask for his opinion. The unequal collaboration seems to frustrate Blue, and apparently, he capitulates: When recording the answer, Blue is no longer providing any back-channeling while Red engages in a monologue of almost four minutes. In contrast, **Dyad 2** shows a high collaboration flow. Only during reading the instructions, there is a short period where none of the subjects talk; but then they engage in a mutual discussion for the rest of the problem solving, and react to each other's questions and proposals.

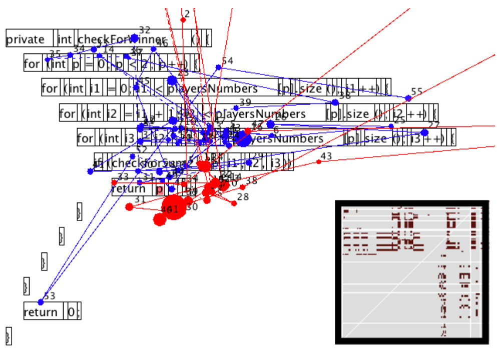
The two cross-recurrence plots that correspond to the 10 minutes of interaction are strikingly different (see Figure 2). For Dyad 1, the density of recurrence points on the diagonal is not very marked. The numerous white stripes are indicative of frequent scrolling. For Dyad 2, clearly defined rectangular areas appear along the diagonal of the plot. This pattern is indicative of gaze coupling, Red and Blue look at the same parts of the code within a few seconds of each other, and hence explore the code together, one criteria for a good collaboration flow. The square patterns are also indicative of some level of stationarity in the exploration of the code (maybe also related to the difficulty of the task), i.e. if the collaborators were constantly changing their focus, only a thin line would be visible on the diagonal. Off-diagonal rectangular patterns show that the dyad iterates through the code by revisiting sections of the code on several occasions.

Figure 2. Gaze Cross-recurrence Plots for Dyad1 and Dyad2. The bold rectangle delimits the moment where the dyads speak about the "check for winner" method and is analyzed below under "mutual understanding" (Full resolution images are available at http://bit.ly/9KQslc and http://bit.ly/aQbhk4).

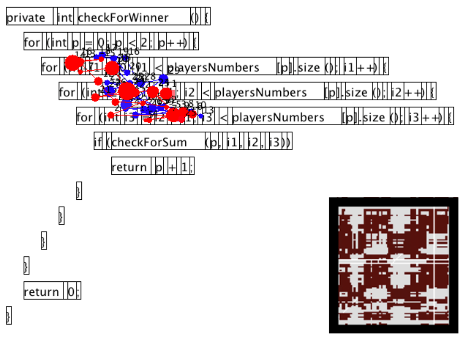
Sustaining Mutual Understanding

Dyad 1 displays difficulties in gaining a common ground. Although Blue sometimes uses verbal references to point out which part of the code he is looking at, the lack of back channeling from Blue indicates that he has difficulties in understanding his partner's comments. In contrast**, Dyad 2** successfully builds a common ground by frequently back-channeling, answering with relevant next turns, and providing verbal or deictic references to draw the partner's attention to specific sections of the code. For instance, towards the beginning of their collaboration, Red suggests, "shall we check the check for winner method?" and thereby guides his partner's attention. Similarly, during their collaboration, Blue highlights a line of the code to ask his partner to explain him the function of the line. However, the relation between selection and mutual understanding is not straightforward. From the dyadic data that we are currently analyzing on 42 dyads we see that there is no correlation between the proportion of interaction time during which subjects select some part of the code and the cross-recurrence. Selection alone is not favoring higher cross-recurrence. Indeed, in Dyad 1, at moments Red frantically selects text in pace with his own reading and speaking activity but without the intention to ground the focus of conversation with his partner.

To illustrate the correspondence between cross-recurrence plots, gaze patterns and dialogue, we compared the interaction of both dyads while they analyze the method "check for winner" in Figure 3. Red who is assisted by Blue only for a vocabulary question dominates the interaction for Dyad 1. While Red selects elements quite often while speaking, Blue does not follow his gaze. In Dyad 2, the two collaborators traverse the code synchronously and produce a socially distributed production started by Red and completed by Blue (Roschelle & Teasley, 1995).



Dyad 1: 56:40 to 57:17
Red: and this part, "check for winner", it actually, p, eh, first look p is for player and those indexes are, eh, for their, their list of taken numbers, and at one moment it simply … eh … gives those indexes and gives what eh what player eh
Blue: are chosen, yes,
Red: yeah thanks, and if that sum is equal to 15 it will simply return p + 1
Blue: hm hm



Dyad 2: 17:49 to 18:05
Red: oh look the initial, the start <select int before i1> of the loop <select int before i2>is different each time <select int before i3>
Red: here
Red: first I1 <select i1> starts at 0 <select 0>
Red: i2 starts at ...
Blue: yeah i2 will start a 1 ... and i3 will start a 3
Red: <select 1 after i2>
Red: yeah so we take the first second and 3rd number and we iterate like that-

Figure 3. Fixation Plot with the Corresponding Cross-recurrence Plot Detail (left) and Transcript for Explanations Given by Dyad1 and Dyad2. Both explanations are about the "check for winner" method.

Technical Coordination

Dyad 1 has a hard time to effectively coordinate the usage of the technical environment. Red is frequently scrolling up and down. The scrolling obviously hampers Blue in his attempts to understand the code: He frequently tries to counter steer by scrolling in the other direction and even explicitly addresses his difficulties "It's just hard to navigate into the program". Still, the difficulties remain, and it is not clear if Red realizes that the screen is shared and scrolling up and down affects both partners. **Dyad 2** successfully coordinates the technical environment. Already at the start of the collaboration, Red addresses the technical coordination by asking his partner if the scrolling also affects his screen. Also in the further progress of the collaboration, they frequently explicitly coordinate the scrolling; for instance, Blue asks his partner if he agrees to scroll further down. Furthermore, the dyad uses the highlighting function to guide their partner's attention. Symmetric white stripes in the cross-recurrence plots indicate scrolling. In Dyad 1, the scrolling stripes are much more dense than in Dyad 2. An indication for effective technical coordination is that gaze is recurrent soon after a scrolling episode, indicating that collaborators did not loose their partners when changing the focus. Because this is not very apparent on the plots, an associated numerical indicator would probably be more informative.

Task Division and Participation Symmetry

The interaction of **Dyad 1** is characterized by a high amount of individual work phases and task division. During the interaction, there are several sequences about a minute and longer where they collaborators do not engage in interaction. The collaboration of **Dyad 2** looks fairly different. There is only a short individual work period at the start of their problem solving where each collaborator reads the instructions; then they engage in mutual interaction. Task division results in low recurrence if it implies that collaborators look at different places of the workspace. Recurrence plots would feature a sequence of off-diagonal elements followed by diagonal recurrence at the end of the interaction if parallel work ended by sharing and reviewing the results. With **Dyad 1** however, the low recurrence seems to stem more from asymmetrical interaction rather than coordinated division of labor. Red basically answers the question on his own; even though Blue first tries to engage in the collaboration, he seems to give up during the further course of the interaction. In contrast, both partners of the **Dyad 2** are equally engaged in the participation, even though with slightly different roles: While Blue frequently asks questions for further clarification, Red provides more explanations. The different roles may be explained by the differences in the partner's prior knowledge.

Gaze recurrence is closely linked with effective dialogue. In **Dyad 1**, sustained audio signal only appears at minute 3 and is afterwards clearly asymmetric in favor of Red. Blue joins in for two minutes between 12 and 14 but then remains silent for the rest of the interaction. Blue only speaks for 4% of the time, whereas Red speaks for 15% of the total time. In **Dyad 2**, the Blue and Red collaborators hold the floor for respectively 41% and 43% of the total time. Of course, the asymmetry of the audio signal is much easier to measure than the level of gaze recurrence and is certainly a simpler indicator for participation asymmetry.

Cross-recurrence in Quantitative Approaches

We have seen the descriptive value of cross-recurrence plots by contrasting two examples of "good" and "bad" interaction quality. In more quantitative approaches to interaction analysis, researchers need a numerical counterpart to the visual patterns that represent the coupling of the collaborators gaze. High recurrence is visible on a cross-recurrence plot where points on the diagonal are dark. More, we are not only interested by the main diagonal (collaborators look at the same target exactly at the same time) but by an "interaction span" during which subjects' gaze follows their partners'. The graph in Figure 4 shows the rate of cross-recurrence as a function of time shifts between the fixations of one subject and the other. Each value on the x-axis corresponds to the time span between the main diagonal on the recurrence plot and a parallel diagonal shifted in time. The values on the y-axis represent the proportion of dark pixels on this diagonal. The cross recurrence is highest in the period of -2 to +2 seconds around the main diagonal, and we used the average value of the cross recurrence rate during this time span as a numerical indicator. The rate could be computed for longer periods of time (e.g. +- 4 and +- 6 seconds). When computed on a larger sample, the resulting indicators are highly correlated with the cross-recurrence rate for a span of +- 2 seconds. Cross-recurrence at much higher time shifts might be indicative of the collaborators revisiting of shared areas of interest (see our discussion above about off-diagonal regions in Figure 1) but does not indicate close coupling anymore.

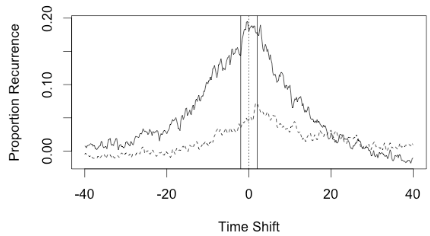


Figure 4. Cross-Recurrence Proportions by Time shifts for Dyad 1 (dashed line) and Dyad 2 (full line).

Conclusion

Contrasting two dyads provided first insights into how gaze cross-recurrence is related to dimension of the quality of collaboration. The most promising result is that high gaze recurrence seems to be typical of a "good" dyad where the flow of interaction is smooth and where partners sustain each other's understanding. It is now necessary to qualify the correspondence between the structure of interaction and graphical patterns in recurrence plots more systematically. One way to do this consists of computing numerical features that reflect graphical structure of the plots (e.g. symmetry, density, gaze transition paths from targets to targets) and to use them to discriminate between categories of interaction quality. An important question for these indicators is also whether they are task-specific and whether they depend on the stimulus displayed on the screen. Normalization of the recurrence measure with regards to a random baseline is required to be able to generalize the approach.

The use of low-level indicators that can be computed in real time would allow to a) automate the collection of evidence to provide real-time feedback, b) to inform and enrich the analysis of interaction by researchers.

Bibliography

Cherubini, M., Nüssli, M.-A., & Dillenbourg, P. (2008) Deixis and gaze in collaborative work at a distance (over a shared map): a computational model to detect misunderstandings. In Proceedings of the International Symposium on Eye Tracking Research & Applications (ETRA2008), p. 173-180. ACM Press, 2008.

Cherubini, M., Nüssli, M-A., & Dillenbourg, P. (2010). Indicating and looking in collaborative work at distance. Journal of Eye Movement Research 3(5):3, 1–205.

Clark, H. H., & Brennan, S. A. (1991). Grounding in Communication. In Resnik, L. B., Levine, J. M. & Teasley, S. D. (Eds.). *Perspectives on socially shared Cognition.* Washington: APA Books.

Dale, R., & Spivey, M. J. (2006). Categorical recurrence analysis of child language. *Language Learning.*

Dale, R., Warlaumont, A. S., & Richardson, D. C. (in press). Nominal cross recurrence as a generalized lag sequential analysis for behavioral streams. *International Journal of Bifurcation and Chaos*.

De Wever, B., Schellens, T., Valcke, M., & Van Keer, H. (2006). Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. *Computers and Education, 46,* 6–28.

Diziol, D., Rummel, N., Kahrimanis, G., Guevara, T., Holz, J., Spada, H., & Fiotakis, G. (2008). Using contrasting cases to better understand the relationship between students' interactions and their learning outcome. In G. Kanselaar, V. Jonker, P.A. Kirschner, & F. Prins (Eds.), *International perspectives of the learning sciences: Cre8ing a learning world. Proceedings of the 8th International Conference of the Learning Sciences (ICLS 2008), Vol 3* (pp. 348-349). ISLS

Eckmann, J.-P., Kamphorst, S.O., & Ruelle, D. (1987) Europhys. Lett. 5, 973.

Firestone, W. A. (1993). Alternative arguments for generalizing from data as applied to qualitative researh. *Educational Researcher, 22*(4), 16-23.

Lord, R. G., & Levy, P. E. (1994). Moving from Cognition to Action: A Control Theory Perspective. *Applied Psychology: An International Review*, *43*(3), 335-398.

Meier, A., Spada, H. & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning, 2*, 63-86.

Nüssli, M.-A., Jermann, P., Sangin, M., & Dillenbourg, P. (2009) Collaboration and abstract representations: towards predictive models based on raw speech and eye-tracking data. In *CSCL '09: Proceedings of the 2009 conference on Computer support for collaborative learning*. ISLS, 2009.

Pietinen, S., Bednarik, R., Glotova, T., Tenhunen, V., & Tukiainen, M. (2008) A method to study visual attention aspects of collaboration: eye-tracking pair programmers simultaneously. In *Proceedings of the 2008 symposium on Eye tracking research & applications (ETRA '08)* (pp. 39-42): ACM, New York, NY, USA.

Richardson, D. C., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science, 29*, 1045–1060.

Richardson, D. C., Dale, R., and Kirkham, N. Z. (2007). The art of conversation is coordination: Common ground and the coupling of eye movements during dialogue.*Psychological Science*, 18(5):407-413.

Roschelle, J., & Teasley S. D. (1995). The construction of shared knowledge in collaborative problem solving. In C. E. O'Malley (Ed.), *Computer-supported collaborative learning*. (pp. 69-197). Berlin: Springer-Verlag

Rummel, N., Deiglmayr, A., Spada, H., & Kahrimanis, G. (2011). Analyzing collaborative interactions across domains and settings: an adaptable rating scheme. In: S.Puntambekar, C. Hmelo-Silver, & G. Erkens (Eds.), *Analyzing interactions in CSCL: Methodologies, approaches and issues*. (pp. 367-390) New York: Springer.

Rummel, N., & Spada, H. (2005). Learning to collaborate: An instructional approach to promoting collaborative problem-solving in computer-mediated settings. *Journal of the Learning Sciences*, *14*(2), 201-241.

Soller, A., Martinez, A., Jermann, P., & Mühlenbrock, M. (2005). From mirroring to guiding: A review of the state of art technology for supporting collaborative learning. International Journal of Artificial Intelligence in Education, 15, 261–290.

Strijbos, J.-W., Martens, R. L., Prins, F. J., & Jochems, W. M. G. (2006). Content analysis: What are they talking about? *Computers & Education, 46*, 29-48.

Suthers, D., Law, N., Lund, K., Rose, C., & Teplovs, C. (2009). Common objects for productive multivocality in analysis. In C. O'Malley, D. Suthers, P. Reimann, & A. Dimitracopoulou (Eds.), *Computer Supported Collaborative Learning Practices - CSCL 2009 Conference Proceedings*, (pp. 225-226). International Society of the Learning Sciences.

Acknowledgements

This research was funded by the Swiss National Science Foundation (grants #K-12K1-117909 and #PZ00P_126611) as well as by a STELLAR incubator grant to the second author, Dejana Mullins.

Training in Virtual Training Environments: Connecting Theory to Practice

Johannes Moskaliuk, University of Tuebingen, Germany, johannes.moskaliuk@uni-tuebingen.de
Johanna Bertram, Knowledge Media Research Center, Tuebingen, Germany, j.bertram@iwm-kmrc.de
Ulrike Cress, Knowledge Media Research Center, Tuebingen, Germany, u.cress@iwm-kmrc.de

Abstract: Many institutions, such as the police, use virtual environments as part of their training and education programs. Typical scenarios that are being trained virtually are of high complexity, require intensive communication between team members, consist of situations in which individual human beings are in danger, and are not trainable in real-life situations. In this paper we describe the virtual training environment *VirtualPolice*. Its theory-based development process is explicated to show how theory was connected to practice. We present which specific needs were evaluated in order to define training goals. We introduce our theoretical base and argue why such virtual training environments can support the acquisition of knowledge-in-use, the ability of team members to take the perspective of others, and reflection of team performance. This makes it possible to identify those features of virtual training environments that enable successful training, and to point out how these may be implemented.

Introduction

Members of institutions such as the police force, fire department or rescue teams are often confronted with scenarios in the course of their work which could not be trained before because of the high costs, the danger, time or effort that such training would have required. Nevertheless, these members of such teams have to react adequately to their daily challenges. They have to act as real teams, solve complex problems, ensure their own safety and avoid dangerous incidences. They need adequate preparation for this type of teamwork that will ensure that they can handle their jobs efficiently, flexibly and safely. Many institutions have recognized the potential of virtual worlds and serious games as tools to train their teams for challenges which could not possibly be trained in reality. This has resulted in various virtual training environments (VTEs) that were developed for different training purposes, and which are often integrated into vocational education and training programs (Rose, Attree, Brooks, Parslow, & Penn, 2000). In the health care sector, for example, the use of virtual realities for training purposes is on the increase (Riva, 2002). VTEs have also been developed to train fire fighters (Julien & Shaw, 2003).

A Virtual Training Environment (VTE) allows people to experience difficult situations without taking any physical risks, and without the need of going to the limits of their physical strength. Any errors made during this training will not lead to negative consequences, and trainers have the flexibility to define different scenarios. Such typical scenarios that may be trained virtually are of high complexity and have no single "correct" solution. They are based on a specialization of tasks within a team, require intensive communication between team members, consist of situations in which individual human beings are in danger, and are not trainable in real life situations.

From the point of view of computer-supported collaborative learning, both theory and practice, these complex training scenarios and virtual training environments are very interesting. The focus here is not only on knowledge, but also on the members' experiences with complex situation that will influence the performance of the entire team. So the aim of a VTE is to combine the different knowledge and skills of the individual team member to a high-performing team. This leads to the three key aspects, which are the base of our goal to connect theory to practice:

(1) *Knowledge-in-use:* Team members need knowledge-in-use to handle complex scenarios. Knowledge-in-use is a combination of different types of knowledge that are necessary to perform a given task, solve a problem, or handle a complex situation.

(2) *Perspective Taking:* Team members have to learn to take the perspective of the others in a complex situation. They have to know about their own skills as well as the skills of other team members, and have to build a shared mental representation of knowledge that exists within in the team in order to respond adequately.

(3) *Team Reflection:* Performance of a team depends on its ability to integrate diverse knowledge and construct a shared understanding of complex situations. Teams need to reflect about own strategies, behaviour and performance. This requires a permanent exchange of one's own experience and the shared experiences within the team.

We argue that VTEs can support the acquisition of knowledge-in-use, the ability of team members to take the perspective of others, and a reflection of individual knowledge and team performance. This paper explicates the theory-based development process of a VTE, as part of our design-based research program. It starts with a description of the specific needs of our target group, a federal state police department in Germany. We introduce our theoretical base and argue why such virtual training environments can support what we call the three key aspects: knowledge-in-use, perspective taking, and team reflection. We introduce the co-evolution model (Cress and Kimmerle, 2008; Cress, Kimmerle, & Held, 2010), which focuses explicitly on the interplay between individual and team learning, and use this model as a framework to integrate our theoretical assumptions. The paper concludes with some features of a successful VTE, connecting theory to practice, and presents the VTE *Virtual Police*, which was developed as part of our work.

Field: Virtual Training for Police Officers

Police officers are confronted with a variety of problems every day. To be prepared for difficult and dangerous situations, they need permanent and constantly updated training, regardless of their level of tenure, education and experience. So the police of the German state of Baden-Wuerttemberg has adopted a broad training and education concept. It is based on an education and knowledge management system that is run by the state Ministry of the Interior, a system which integrates and coordinates different educational provisions, ensuring a continuous improvement of the skills of police officers in the state. This knowledge-management platform combines conventional education with training sessions and e-learning, and the virtual training environment *VirtualPolice* is, so to speak, an extension of this platform.

VirtualPolice makes it possible to train operations, which were previously untrainable, such as the interaction between officers on the ground with a helicopter crew. Due to the specific requirements of using helicopters (for example, availability, organizational effort, keeping down costs and noise), such operations can only be trained restrictively. So far, only so-called Special Forces have been trained intensively for such cases. But all police officers (22,000 in the state) might be confronted with a helicopter call, so such training is relevant to all of them.

The development of *VirtualPolice* is part of a design-based research program, and based on strong cooperation between partners with different backgrounds. The technical development of *VirtualPolice* is carried out by TriCat, a software company specialized on programming complex virtual environments. Police trainers from a local training and competence center and the helicopter crew are involved in the development of *VirtualPolice* as domain experts. Our part within the project is to accompany the development of *VirtualPolice* as researchers and connect theory to practice.

The first step of our development process and goal to integrate theory and practice was to become acquainted with the specific training needs of our field partners, the police force of Baden-Wuerttemberg. Due to the high complexity of the tasks that were to be trained and in the absence of tested and standardized work routines, we decided to use non-standardized interviews and expert workshops to identify current problems and training goals. Officers, police trainers and members of helicopter crews were interviewed about their experiences, and corresponding operations were staged and analyzed. Subsequently, specific training needs were identified, based on the outcome of the analysis, which we had conducted. These training goals may be assigned to the three key aspects, which are listed below.

Knowledge-in-use: Specific courses of action, which are required in dangerous operations, need to be trained. Police officers have to react adequately and fast in every situation. A sense of danger should evolve and be retained in every situation. At the same time, the presence of the helicopter leads to greater stress, due to the noise produced by the helicopter, to which the police officers are not adapted. Their teams have to decide under enormous pressure, even if the available information is sparse.

Perspective Taking: Many problems resulting from interaction between officers on the ground with a helicopter crew are due to a lack of knowledge of what a helicopter and its crew are really able to do. Police officers may, for example, misjudge what the helicopter crew can see from above. In addition, both the team on the ground and the helicopter crew need to adjust their communication style to the situation. The helicopter crew will, for example, need to be informed about intended actions on the ground.

Team Reflection: One central need is to make individual experiences available to other team members. If team members have finished an operation, it should be ensured that problems that occurred and their successful solution are communicated to the other members. This should lead to individual learning and, at the same time, a higher performance of the whole team. The same is true of the integration of experiences of helicopter crews into the training of ground forces. Integrating such different experiences makes it possible that the entire team is able to reflect on their knowledge and team performance. As helicopter crews, so far, have not been trained together with officers operating on the ground, the opportunity of training together will improve learning at an organizational level.

The next step of our design-based research program is to find theories that provide insights in how these training goals may be achieved.

Theory: Knowledge-in-use, Perspective-taking, Team Reflection

Current constructivist approaches assume that active construction of knowledge required active collaboration, using real or digital objects as epistemic artifacts. This has led to a plethora of theories that deal with situations in which groups use some shared digital artifact to construct knowledge. The knowledge building concept (Scardamalia & Bereiter, 1994) addresses how a community of learners will manage to develop knowledge jointly. This approach focuses, to a large extent, on using computer technologies to form a knowledge building community. A knowledge building community acts like a scientific community, with the aim of constructing new knowledge and developing some theory to understand the environment. The knowledge creation spiral (Nonaka, 1994; Nonaka & Toyama, 2003) mainly focuses on the building and transfer of tacit knowledge (Polanyi, 1966). Most knowledge is contained in the experience of individuals, and can only be expressed verbally and transferred to others with some difficulty. The authors describe four processes, which dynamically build on each other: socialization, externalization, combination, and internalization. The knowledge building model of Stahl (2000) describes how social knowledge building and personal understanding take places and influence each other. It also highlights how cultural artifacts that are used in activity influence personal comprehension and learning. The knowledge creation metaphor (Paavola & Hakkarainen, 2005) focuses on a mediated process of knowledge creation, where shared objects of activity are developed jointly. The mutual creation of new material and conceptual artifacts leads to an advancement of knowledge, discovery and innovation. The idea of knowledge maturing (Schoefegger, Weber, Lindstaedt, & Ley, 2009) describes how knowledge matures in the course of time from expressing individual ideas, collaboration on shared artifacts, standardization in documents, to a creation of learning objects and standardized training programs. During this process, knowledge will develop, be more and more integrated in an organization, and shared between all its members

These theories, which were outlined here very briefly, all share the idea that the development of knowledge takes place within communities and that epistemic objects (digital or real world objects) will support this development of shared understanding and community knowledge. These theories are also a good base to understand how training in a virtual environment leads to a connection of different knowledge and skills of single team member to achieve a high-performing team. To connect theory to practice in our present context, we focus on the three key aspects knowledge-in-use, perspective taking, and team reflection. These are the basis of the features of the VTE, which we developed.

Knowledge-in-use

Theoretical assumptions: To solve complex situations and difficult tasks, the team as whole and its individual members need a great deal of experience from similar situations. This experience results from practice and well-tested sequences of action during team members' work. Knowledge of the type that is meant here may be described as knowledge-in-use (De Jong & Ferguson-Hessler, 1996). Knowledge-in-use is a combination of different types of knowledge that are necessary to perform a given task, solve a problem, or handle a complex situation. Knowledge-in-use may be defined as knowledge about activities or tasks that are performed frequently and therefore well established in the action routine of a person. Knowledge-in-use is embedded in daily challenges and in most cases implicit, because it is based on experienced work routines, which are often carried out unconsciously (Smith, 2001). Knowledge-in-use is highly *situated* (Greeno, 1998), as it is tied to specific contexts, situations or circumstances. Individuals build *situational knowledge*, that is, a relation between a situation that requires certain knowledge and the knowledge itself, about "situations as they typically appear in a particular domain" (De Jong & Ferguson-Hessler, 1996). This relation helps to identify relevant features of a current problem, to build an adequate representation of the problem, and to retrieve additional (declarative, conceptual or procedural) knowledge to solve a problem. In contrast to declarative or conceptual knowledge, it is not easy to transfer knowledge-in-use from one person to another. The larger part of knowledge-in-use consists of implicit knowledge about sequences of action (Polanyi, 1966). Members of an organization have to be aware of their individual work routines and experiences. They have to draw general conclusions from situated knowledge-in-use, which can then be presented as abstract knowledge and transferred to other situations and contexts. Others than have to internalize this transferred knowledge and integrate with own experience in order to make it adaptive.

Adoption to VTE: VTEs are a good means to support the development of knowledge-in-use. The VTE is a tool to simulate real situations. This should help a trainee to draw the relation between a specific situation and the relevant knowledge-in-use, which is necessary to respond adequately. This will not only lead to an appropriate reaction in the trained situation, but also allow a transfer of knowledge-in-use to similar real situations. The observation of what other members of a team are doing will also lead to a form of learning through socialization (Nonaka, 1994). Experiences are shared and tacit knowledge can be communicated from one person to another only through direct experience. So experiences can be transferred through observation and imitation, and, as a result of this transfer, the observer acquires knowledge, but this will also remain tacit.

A VTE provides opportunities, close to real experience, to acquire new knowledge-in-use by observing other users. Team members can observe strategies and actions of other members and learn from their success or failure. Even if trainees are not part of a current scenario themselves, they will benefit from their position as outside observers. They can adopt the observed strategies as part of their own knowledge, and give feedback to others from a third-party position. It should be emphasized here that learning is a socio-cognitive process. Different strategies of team members to handle a specific situation or to solve a problem may conflict with each other. To become a well functioning team, they have to find a common understanding of the situation and develop adequate strategies. A VTE can serve as a cognitive tool to highlight different strategies and provide possible alternative solutions to conflicts, which may occur by training together. In this way, a learning team will be formed in which all members with different backgrounds and experiences can bring in their own knowledge to the benefit of all.

Summary: Knowledge-in-use is an essential outcome of this form of training. Team members have to acquire knowledge-in-use to integrate different types of knowledge and solve a given task. A VTE should support this acquisition of knowledge-in-use.

Perspective Taking

Theoretical assumptions: The concept of perspective-taking describes the "process of imagining the world from another's vantage point or imagining oneself in another's shoes" (p.110) (Galinsky, Ku, & Wang, 2005). From a social psychology point of view, perspective-taking results in an overlap of a mental representations of the self and the other and leads to a decreasing of stereotype expression and in-group favouritism (Galinsky & Moskowitz, 2000). What is relevant in the context of knowledge-in-use acquisition is that perspective-taking will also influence interaction and coordination within a team. Perspective-taking leads to higher similarity of the group members' behavioural strategies and fosters the coordination of social behaviour (Galinsky et al., 2005). A team member has to build a mental representation of his or her own perceptions, strategies and competences and those of the others. The team as whole will then develop a shared understanding of complex situations and strategies to act as a team. This shared understanding does only include knowledge about the others' expertise (cf. the concept of a transactive memory system (Wegner, 1986)) but also knowledge about how to act together and handle a complex situation. This team knowledge is an important part of knowledge-in-use that is required to solve a complex task. Taking the perspective of other team members will also help to prevent communication failures. Members who know about different vantage points within a team (e.g. the different perspectives of helicopter and ground crews members) are able to improve their interaction strategies in an adequate way. They adapt their communication to the perceived knowledge of the other members (Krauss & Fussell, 1991). This supports the common ground (Clark, 1996) within a team and is an important condition of success even in hazardous situations.

Adoption to VTE: We argue that a VTE can support perspective-taking within a team by simulating the perspective of other members. A team member is in a position here to experience the virtual world from the perspective of another member who has different tasks and commissions. The VTE enables the team easily to swap positions, equipment or viewpoints in order to build an understanding of the others' perspectives and of the whole situation. As a result, the team will build a shared mental representation, which can be used in similar situations in real-world settings. Moreover, VTEs provide the opportunity to review finished virtual training sessions several times from different perspectives. One team member can experience the training session from the vantage point of another member. Then a trainee builds a mental representation of the whole situation and integrates the different perspectives and tasks of single team members into his or her own knowledge-in-use.

Summary: The ability to take the perspective of others is an essential outcome of successful training. Team members have built a shared mental representation to perform as a team. A VTE should support perspective-taking by providing the opportunity to switch between different perspectives.

Team Reflection

Theoretical assumptions: There are two relevant outcomes of training: Individual learning and team performance. Both outcomes influence each other because integration of different experiences leads to reflection on one's own knowledge, and enables a development of the whole team. The VTE acts as an epistemic artifact that can visualize the actions of the team members and support reflection processes. This mutual evolution of the individual and the team is described by the co-evolution model, as proposed by Cress & Kimmerle (2008). This model refers to the work of Scardamalia and Bereiter (2003, 2006), but points out, at the same time, that individual learning and collective knowledge building are two parallel and equally important processes, which support each other. Individual learning describes the internal processes that take place during collaboration and lead to better understanding of the environment and greater conceptual knowledge. Knowledge building, on the other hand, as a collective creation of public knowledge, is regarded as an improvement of ideas or the development of new ideas, not as a search for a perfect or true solution. It is as a discourse-oriented process in the sense of joint problem solving. Shared digital artifacts can support knowledge building, as they enable participants to contribute their own theories, models, examples, visualizations, notes, and other epistemic artifacts. The shared digital artifact initiates a dynamic and self-organized process, in which ideas are formulated, discussed, revised, or rejected. The co-evolution model regards the exchange of knowledge as an interplay between individuals and a community.

Adoption to VTE: The co-evolution model offers a framework to describe the mutual development of individual knowledge and team performance during virtual training. The co-evolution model does not explicitly focus on knowledge-in-use, but knowledge building will, of course, also take place if learners handle authentic real-life problems that require knowledge-in-use. A VTE can support this co-evolution of the members' individual knowledge and team performance by providing a platform for mutual action. This form of joint training within a VTE leads to a permanent exchange of individual and shared experiences within a team. This leads to an integration of diverse knowledge, and constructs a shared understanding of complex situations. Unlike real-life training, it is possible in a VTE to record a complete training scenario and replay it later from different perspectives. Team members can review and observe their own behavior and identify failures or potential for improvement. The VTE provides a form of objective feedback on one's own reactions and strategy, for example, by measuring reaction times, replaying communication behavior or visualizing strategic components. This supports and validates the subjective feedback of trainers or other team members, and will certainly improve individual development. The team performance will also be developed further through such mutual training and the possibility to review training scenarios. This can help to detect communication failures, misconceptions or conflicts during the collaboration. Deficits can be identified and different viewpoints integrated into shared team knowledge. This also leads to learning at an organizational level, because the different experiences that were made during virtual training scenarios can be stored and retrieved later. Thus, experiences made by other teams are available to others, who can review their own behavior and connect it to their own knowledge-in-use. This facilitates the integration of different experiences and the collective creation of knowledge-in-use at the organizational level.

Summary: The mutual development of individual knowledge and team performance is an essential outcome of training. Team members reflect about their own behaviour and develop individual knowledge, at the same time team performance will evolve during the virtual training. A VTE should support this co-evolution by providing the opportunity to switch between different perspectives, record and replay finished scenarios.

Connecting Theory to Practice

Coming back to the specific needs of our partner from the field, we can connect theory to practice and formulate those features of a VTE, which provide ideal conditions for training complex and dangerous scenarios. We use the three key aspects to structure these features, bearing in mind most of them are relevant to more than one key aspect.

Knowledge-in-use: A VTE supports the acquisition of knowledge-in-use by providing a flexible and reviewable training space.

(1) Training is possible even if training in reality would be too dangerous or too expensive.
(2) Specific courses of action demanded by typical operations may be trained many times and with different team configurations, in order to acquire some routine and flexible reactions.
(3) The difficulty of a task may be modified, for example, by starting with a reduced scenario and increasing its complexity.
(4) The VTE can provide additional scaffolds to support trainees at an early stage of training, and this support can be faded out later.
(5) Training may pause at any time, to reflect on the current status, define possible problems and re-adapt a strategy.

Perspective Taking: A VTE supports perspective taking within a team by simulating the perspective of other members.

(1) A VTE can simulate the vantage point from which other team members view the situation.
(2) Team Members can see their own behavior from the perspective of others, e.g. by seeing the helicopter crew's view of the ground.
(3) Team members can take the position of an outside observer, looking at scenarios from a wider perspective. This supports the acquisition of strategic knowledge.

Team reflection: A VTE supports team reflection, which will enhance individual learning and team performance.

(1) Completed training scenarios are stored and can be retrieved later.
(2) Team members can learn from others' experiences by reviewing them later.
(3) Recorded communication, e.g. between helicopter crews and ground forces, may be analyzed to detect communication failures.
(4) Data on the performance of team members can be aggregated to visualize, for example, navigation routes and resulting dangerous situations.

The VTE *VirtualPolice*

The virtual training environment *VirtualPolice* was constructed to train both ground forces and helicopter crews, although the integration of fire fighters, emergency medical services or other emergency services in a police operation can also be implemented, making this VTE highly adaptable. *VirtualPolice* simulates 150 square kilometres of accessible terrain, with villages and towns characterized by individual townscapes (see figure 1). Some building structures are accessible, and open areas in the countryside and woodlands exist as well. Day and nighttime in the simulation can be influenced and weather conditions be set, so that training scenarios can be adjusted to specific training needs. The trainees are represented by avatars, dressed either like police officers or suspects. Each participant perceives what is going on visually and with sounds through his or her own ego perspective, and can move freely through the VTE. Furthermore, the scenario can be populated by autonomous agents, guided by artificial intelligence. In addition, police patrol cars and normal vehicles are available.

Figure 1. Screenshot of *VirtualPolice.*

The available helicopter is represented with realistic pictures and sounds, as seen by the participating ground forces, and can be navigated either by trainers or helicopter pilots. A helicopter pilot in the simulation can choose between natural day/night view, use image-intensifying glasses at night and an infrared camera in order to provide a heat image of the scenario (figure 2).

Figure 2. Helicopter Perspective within *VirtualPolice* Using the Infrared Camera.

The training participants use headsets to communicate through spoken language. Trainees, trainer and helicopter pilots can use a simulated two-way-radio, and when they are close enough to each other in the VTE, they can hear each other without the need to use this transmission. This simulates real-world conditions in which the ground force team talk to each other face-to-face, but have to communicate by radio with crews further way or with the helicopter pilot. The trainer's monitor can show any position in the scenario and switch between different trainees' views and the helicopter view. He or she can communicate with the simulated police officers, with the suspects, or with all of the participants at the same time. This allows the trainer to give additional support, providing scaffolds or interrupt the training. The trainer can re-arrange the training scenario by placing any participant, the helicopter or the suspect into any position within the VTE and re-start the training.

VirtualPolice is designed to train up to 12 participants simultaneously, and the system requirements of the software are low, so it will run on standard computers with no need for further equipment - only a keyboard, mouse and headsets are required. The VTE is part of the organizational network of the state police and can be used online. The trainees can remain in their local police stations or departments, but train together with other teams. *Virtual Police* provides the option to record the entire training scenario and replay it later from different perspectives. The trainer can set markers and written comments during the training in order to highlight relevant situations, for example, communication failures or dangerous behavior of the trainees. These markers can used to provide a structured feedback after the virtual training.

Application of the Training

The training concept of *VirtualPolice* consists of a preparatory stage with theoretical information, a practical stage using the VTE, and a feedback phase, which also integratesthe VTE. During preparation, the trainer first explains to the participants the content and the importance of this training. To underline this point, videos of real operations with dangerous developments are introduced and discussed. Then, information on appropriate behavior during such operations is provided and discussed. Finally, the training goals are defined explicitly, pointing out the focus of the practical stage. In this stage the participants are first instructed how they have to to use the VTE and how to navigate within it, and are given some time to acquaint themselves with the VTE. Subsequently, the prepared training scenarios are executed. The trainees are seated in single cabins in front of a 24-inch monitor and wear a headset with integrated microphone, as an attempt to isolate them as widely as possible from environmental influences (figure 3).

Figure 3. Police Officer Trains Using *VirtualPolice.*

After each scenario, a feedback phase is conducted. The scenario is discussed and every action can be replayed from different perspectives, allowing all participants to take the others' perspectives and to see their own actions from others vantage points. After the practical stage the participants receive feedback from their trainer depending on their performance and progress during the training. Newly acquired skills and knowledge-in-use can be discussed and real-world applications can be illustrated.

Summary

This paper explains our theory-driven approach to the development of a VTE for police officers, focussing on knowledge-in-use, perspective taking and team reflection. We consider these as key factors for efficient virtual training, and try to highlight the importance of sound theoretical foundations and assessments when applying VTEs. But the theoretical basics should not only be kept in mind when designing and developing a training environment. There should also exist, at least in the early stage of implementation, a feedback loop between trainers, trainees, psychologists and software developers. In our example *VirtualPolice,* we conducted after each newly released version of the VTE a training stage to test and explore the training environment, and we also arrange regular workshops involving police trainers, helicopter crew members, ground forces, psychologists and software developers, to discuss future improvements of the VTE. During our training and workshop sessions we discovered, for example, the necessity to have personalized avatars and identifiable streets and city areas. With such feedback loops, we intend so secure a high standard of training, permanently refresh our assessment of training needs, encourage communication between trainers, ground forces and helicopter crews, and connect theory to practice.

References

Clark, H.H. (1996). *Using language*. Cambridge: Cambridge University Press.

Cress, U., & Kimmerle, J. (2008). A systemic and cognitive view on collaborative knowledge building with wikis. *International Journal of Computer-Supported Collaborative Learning*, *3*(2), 105-122.

Cress, U., Kimmerle, J., & Held, C. (2010). The interplay between individual and collective knowledge: Technologies for organisational learning and knowledge building. *Knowledge Management Research & Practice*, *8*, 33-44.

De Jong, T., & Ferguson-Hessler, M.G.M. (1996). Types and qualities of knowledge. *Educational psychologist*, *31*(2), 105-113.

Galinsky, A.D., & Moskowitz, G.B. (2000). Perspective-taking: Decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of Personality and Social Psychology*, *78*(4), 708-724.

Galinsky, A., Ku, G., & Wang, C.S. (2005). Perspective-taking and self-other overlap: Fostering social bonds and facilitating social coordination. *Group Processes & Intergroup Relations*, *8*(2), 109-124.

Greeno, J.G. (1998). The situativity of knowing, learning, and research. *American Psychologist*, *53*(1), 5-26.

Julien, T. U. S., & Shaw, C. D. (2003). Firefighter command training virtual environment. In *Proceedings of the 2003 conference on Diversity in computing* (pp. 30-33). Atlanta, Georgia, USA: ACM.

Krauss, R.M., & Fussell, S.R. (1991). Perspective-taking in communication: Representations of others' knowledge in reference. *Social Cognition*, *9*(1), 2-24.

Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization science*, *5*(1), 14-37.

Nonaka, I., & Toyama, R. (2003). The knowledge-creating theory revisited: knowledge creation as a synthesizing process. *Knowledge Management Research & Practice*, *1*(1), 2-10.

Paavola, S., & Hakkarainen, K. (2005). The Knowledge Creation Metaphor – An Emergent Epistemological Approach to Learning. *Science & Education*, *14*(6), 535-557.

Polanyi, M. (1966). *The Tacit Dimension*. Garden City, NY: Doubleday.

Riva, G. (2002). Virtual reality for health care: the status of research. *Cyberpsychology & Behavior: The Impact Of The Internet, Multimedia And Virtual Reality On Behavior And Society*, *5*(3), 219-225.

Rose, F.D., Attree, E.A., Brooks, B.M., Parslow, D.M., & Penn, P.R. (2000). Training in virtual environments: transfer to real world tasks and equivalence to real task training. *Ergonomics*, *43*(4), 494-511.

Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *The Journal of the Learning Sciences*, *3*(3), 265-283.

Scardamalia, M., & Bereiter, C. (2003). Knowledge Building. In *Encyclopedia of education* (pp. 1370-1373). New York: Macimillan Reference.

Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 97-115). New York: Cambridge University Press.

Schoefegger, K., Weber, N., Lindstaedt, S., & Ley, T. (2009). Knowledge Maturing Services: Supporting Knowledge Maturing in Organisational Environments. In D. Karagiannis, & Z. Jinpeng, (Eds.), *Knowledge Science, Engineering and Management: Third International Conference, KSEM 2009* (pp. 370-381). Berlin: Springer.

Smith, E.A. (2001). The role of tacit and explicit knowledge in the workplace. *Journal of Knowledge Management*, *5*(4), 311-321.

Stahl, G. (2000). A model of collaborative knowledge-building. In *Proceedings of ICLS 2000* (pp. 70-77). Ann Arbor, Michigan, USA: University of Michigan.

Wegner, D. (1986). Transactive memory: A contemporary analysis of the group mind. In B. Mullen, & G. Goethals (Eds.), *Theories of group behavior* (pp. 185-208). New York: Springer.

Acknowledgments

The research presented here is part of a cooperation project of the Knowledge Media Research Center and the University Tuebingen together with the *Polizei Baden-Wuerttemberg* (federal state police department of Baden-Wuerttemberg), the *MFG Baden-Wuerttemberg* (centre of excellence for IT, media, and film), the *TriCAT GmbH* and *T-Systems*.

Analyzing Discourse Synthesis – Use of Semantic Information for Collaborative Writing in Secondary Education

Crina Damşa, InterMedia, University of Oslo, Postboks 1161 Blindern 0318 Oslo, Norway,
crina.damsa@intermedia.uio.no
Gijsbert Erkens, Jeroen Janssen, Research Centre Learning in Interaction, Utrecht University, the Netherlands,
Email: G.Erkens@uu,nl, J.J.H.M.Janssen@uu.nl

Abstract: Discourse synthesis refers to writing activities that involve synthesizing information from multiple textual sources for writing a new text. This study investigates the way groups of secondary education students use semantic information from multiple sources, while undertaking a collaborative, computer-supported writing history task. We analyzed the products of 10 groups (32 students) and the chat discussions of the group members, and we assessed the quality of the group products. The results of the study offer an insight into how students select and integrate semantic information. Most noticeable results show that groups synthesized the information selected from sources based on ideas rather than sources and created new conceptual connections among these ideas. Relationships were found between the quantity and type of information and the products quality, while weak or no relationships were found between type and frequency of chat talk with the degree of integration in the group products.

Introduction

Discourse synthesis refers to reading and writing activities within a specific type of writing task, which requires subjects to synthesize information from multiple textual sources (see Spivey, 1995). To a certain extent, discourse synthesis is employed at all educational levels, from primary school to university and in all disciplines. The products of this activity can vary from reports to essays or literature studies. Such tasks are often used for evaluating students' capacity to synthesize information from different sources and to build an argument. The educational practice shows that the evaluation of this type of written tasks is mainly realized by only considering the final text, and very rarely based on the process that took place in order to produce that text. Nevertheless, there is more about discourse synthesis tasks than this final text. The cognitive and meta-cognitive processes taking place while performing such tasks are of importance. Research studies show that a series of cognitive activities differentiate this type of task from other writing tasks and often make the difference between a good and a mediocre composition (Segev-Miller, 1997). While many studies of discourse synthesis focus more on the product and its characteristics (content, composition), the more recent ones attempt to gain insight into the cognitive and meta-cognitive landscape of the discourse synthesis process. This research study focuses on one specific aspect of the discourse synthesis process, i.e., the way students use the semantic information from the multiple textual sources.

Discourse Synthesis

Studies on discourse synthesis tasks have been conducted with two purposes: firstly, to identify the cognitive and meta-cognitive mechanisms of the process and secondly, to identify the differences in performance between successful and unsuccessful synthesizers, and the factors that might cause these differences. However, strategies identified by the mentioned studies are not unique to the discourse synthesis process. Many strategies identified in the synthesizing process have been found also in general writing task, such as: selecting (Kants, 1998; McGinley, 1992; Higgins, 1992; Spivey & King, 1989), planning (Kantz,1989, McGinley, 1992); elaborating (Higgins, 1992), organizing (Higgins, 1992; McGinley, 1992; Kantz, 1989); evaluating, (Higgins, 1992; Segev-Miller, in press), revising (Kantz, 1989; Higgins, 1992; McGinley, 1992).

Some analyses also provided an insight into the characteristics of the discourse synthesis tasks. A number of studies concerned with the cognition within the writing tasks pointed at the fact that synthesizing information from sources occurs both at textual and conceptual level. Spivey (1997) calls the cognitive processes in discourse synthesis processes *textual transformations*. The production of the new text happens by cognitive means which are: *selecting, organizing and connecting transformations.* Segev-Miller (1997; in press) tried to gain a more comprehensive view on the cognitive processes underlying the performance of this task at academic level. The element of novelty in Segev-Miller's approach is the category of *transforming or intertextual strategies*. These can be briefly described as cognitive strategies that subjects use in order to synthesize information provided by more textual sources and use it for contracting a new idea, phrase or text. Based on her studies, Segev-Miller (1997) categorizes transformations in *conceptual, rhetorical and linguistic*.

We explore first activities at the conceptual level. The activity of *selecting i*nformation occurs mainly during reading, in the context of meaning making. Spivey (1997) maintains that, during reading, meaning is constructed and then text is produced in order to 'signal' that meaning to others. She considers it important to distinguish between two products of discourse processes: mental products and textual products. The mental product is actually the meaning the individual is constructing using the reading text together with prior knowledge as input. The writer, actively producing the mental meaning, is also producing a written product, meant to be read by the others. According to Spivey and King (1989), selecting is guided by a very clear goal that the reader sets, which is to find relevant input for the own text. It is not possible or necessary for the reader to use all information from the given sources. A number of criteria are identified based on which readers select information (Spivey, 1995). These criteria are: *textual* relevance, which refers to the hierarchical placement of content within the whole text; *intertextual* relevance refers to the information occurring in more than one source and again, the placement of the content within the various text sources; the *contextual* criterion involves that writers choose the information that is most relevant to the type of message being written; the *rhetorical relevance* refers to the fact that writers use certain information from sources for particular audiences, and adapt their selection process to the type of audience.

Although named differently, comparable conceptual activities have been identified by Spivey and Segev-Miller in their studies. Spivey (1997) identifies more types of connecting transformations: the *topical connections* are made within texts as well as across texts and have the purpose to maintain the coherence of the discourse clause by clause; the *whole-text connections* are transformations that occur at a lager scale, and are usually intertextual; *nonlinear* connections are comparable with the connections made while using hypertext. Segev-Miller (1997) uses the term c*onceptual transforming,* which consists of deliberate intertextual processing of the source texts in order to identify conceptual connections between them, and the creation of macropropositions, in order to connect the proposition from the sources. When exploring activities at *rhetorical* level, Higgins (1992) identifies activities such as structuring, shaping and re-shaping materials from the various source texts. The connection between the separate semantic entities is realized by means of manipulating propositions originating from the source, such as arranging text into high- and low level propositions, discovering relations between ideas, or looking for super-ordinate categories in order to subsume items. Spivey and King (1989) consider organizing as supplying the content with a new structure. Writers are busy creating different semantic chunks which are fit into the organizational patterns of the original text, but mostly are being used to generate a new kind of organizational pattern (Spivey, 1995). In this type of organizational strategy the writer provided an integrated presentation of the material that has been selected from different sources. Segev-Miller (1997) identified rhetorical activities, which serve to translate the conceptual connections to text. Five rhetorical strategies were identified and classified, based on the degree of cognitive demands every strategy requires. *Summarizing one source text is* a less demanding strategy. *Listing* the source text, compared with Bereiters and Scardamalia's (1987) knowledge-telling, is a strategy which consists of summarizing every source text and presenting these summaries after each other . *Incorporating source texts in one source text* is applied by using one text as a frame for incorporating information; and by adding information from other texts order to replace deleted information from the frame text. The strategy of *decomposing and recomposing* source texts is considered as being of intermediate difficulty, and consists of `synthesizing by ideas rather than by authors` (Segev-Miller, 1997, p. 22). It involves breaking one sources text into propositions, and then looking across the other source texts for propositions that can be related to these propositions from the first source. Finally, *synthesizing the source texts* involves more than two source texts, requires a high level of intertextual processing and is considered the most difficult rhetorical activity.

Finally, Segev-Miller (1997) identifies *linguistic strategies,* used in order to compensate for the low degree or lack of conceptual transforming and intertextual processing. Examples of such activities are *speech acts,* which consists of using certain type of semantic means (usually words) in order to create the impression of connection between the different ideas that are presented; or *lexical repetition*, such as paraphrasing or repetition, used to increase the degree of cohesion at thematic level in their texts.

A Model of Writing Based on Multiple Sources

We constructed a model of text writing, which attempts to emphasize the way information from multiple sources is used for writing a new text. A first phase in the process includes *the selection of relevant information*. In previous studies, little was said about the quantity and type of the selected information, or the choice for certain sources during selection. *Integration of information* is another phase. In this phase, subjects try to create a conceptually meaningful representation based on the selected information, in order to compose the new text. In addition, the research studies mentioned activities at meta-cognitive level, such as planning, monitoring and evaluating, which are meant to offer the guidance and support for the actual writing in process. The recurring assertion of researchers regarding the cognitive activities in writing is that this process of writing is not just a rigid sequence of processes and sub-processes, but more a set of activities organized by the writers in a certain order, depending upon intermediate results of the process. The result of these activities is visible in the intermediate versions of the text (drafts). The way students deal with information from sources during these phases, the type of information they use from sources and how they manage their writing process will be reflected in the final product.

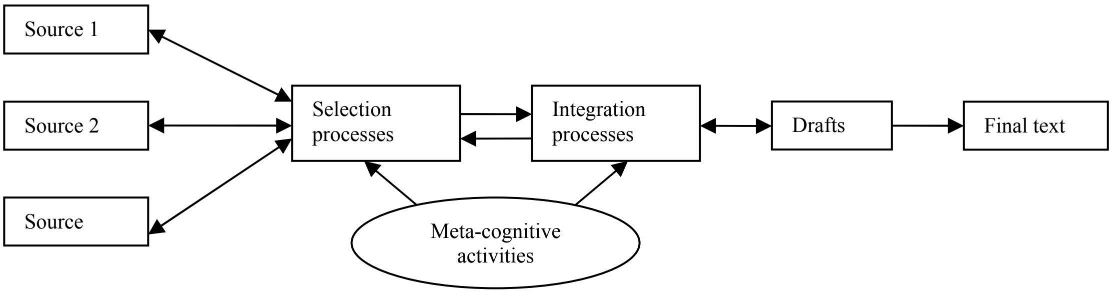


Figure 1. Model of Writing from Multiple Sources.

In relation to the above, the following research questions will be answered: *1. How do groups of secondary school students make use of semantic information from multiple sources while undertaking a computer supported inquiry history task?* and *2. What meta-cognitive activities are reflected in chat discussions of groups of secondary school students, when undertaking a computer supported discourse synthesis history task?* We expect the 'quality' of written text to correlate with the amount of information that students use from the given sources; in other words, students that select and integrate more information from sources will write better texts. Another expectation is that the quality of written text would correlate with the type of information that students use from sources; in other words, students that use information from sources that are considered essential will write better texts. In addition, we expect that there will be a relationship between the type (on- or off-task) and frequency of chat talk with the degree of integration in the final text product.

Methods

The participants in this research project were 32 (mean age of 16 years, SD = .58) eleventh-grade students, along with their history teachers, from a secondary school in the Netherlands. Participating students were randomly assigned to a group of three or four by the researchers, in order to obtain heterogeneous groups with respect to gender and ability. During the experiment, students collaborated in the *Virtual Collaborative Research Institute (VCRI)* designed by Erkens, Janssen, Broeken & Jaspers (2002). VCRI is a groupware program designed to support groups of students during collaborative projects or inquiry tasks. A number of tools, such as *Chat, Sources* and *Co-writer,* were used in order to perform this collaborative writing task, and other tools e.g., the *Statusbar*, *Forum, Planner*, were employed to organize the process.

Students worked in collaboration on a history inquiry task with the title 'Witches and the prosecution of witches'. Based on the six given information sources (and on additional sources, found independently) students had to perform seven assignments, by answering different questions with regard to the topic and write their answers as short descriptive texts, or argumentative essays. Students worked together in groups of 3 or 4, using the VCRI environment to collaborate. In VCRI, groups had access to the sources, could make notes, drafts, communicate with the other members of the group, etc. Through the Co-Writer every member of the group could see how the draft (shared text) evolved to a final version, and the other's contribution to the text. The result of their work was an argumentative text of 700 words about witch prosecution. The sources used for this assignment were original historical texts or translations of original texts about witch prosecution in the 16th and 17th century. The student groups worked the inquiry group task for a period of four weeks. In the first lesson, the teachers introduced the task and gave the necessary instruction regarding to the task, and the most important features of the CSCL-environment were explained. Seven more history lessons were devoted to the inquiry group task. Students were allowed to work on the inquiry task outside normal classrooms hours.

In order to answer the first research question, two types of analyses were conducted. First, we conducted a *semantic text analysis* of the text sources, of the drafts and of the final group products. For this purpose, a parsing procedure was developed based on the theory of semantic representation of Kintsch and Van Dijk (1978). Two types of parsing rules were applied: the macro rules defined by the semantic representation theory (generalization and construction), and the second set of rhetorical rules, developed in order to determine the morphological structure of the propositions. A first aspect treated by these rules is the structure of the (semantic) propositions and the use of grammatical aspects. According to these rules the standard structure of a proposition is: *subject, verb, object, specification of the object*. The subject and the verb are compulsory elements of the unit. Some examples of these rhetorical rules are: the nouns are always singular, with the exception of the situation when a generalization is necessary; or verbs are always used with their infinitive form, present tense and passive voice. This is an example of text: *"Amongst the women was one of unbelievable beauty who was telling all kinds of fabulous stories about what she had done; amongst others, she told she could fly…"* (Translation from Dutch from *Batavische arcadia,* by Johan van Heemskerk), which can be parsed as follows*: Woman is beautiful, Woman says (woman can fly)*. We constructed two different lists of words and their synonyms, used on sources or used in the text assignments. One of the synonyms was chosen and consistently used in all situations. Examples are: devil, was used for *devil, spirits, demons*; illusion(s) was used for *hallucination, visions;* women was used for *girls, women.* The second list regarded the syntactical components of the propositions. Examples from this list (verbs list): accept was used for *to confess, to accept;* accuse was used for *to officially accuse, to call for blame, to give the blame.*

Once sources and group products were parsed, we analyzed the way groups selected and integrated this semantic material. The following aspects were considered: 1) The *quantity of semantic information* selected from sources and used in the drafts and final group products; 2) The *type of semantic information* - from core of secondary sources; 3) The *origin of the selected information* (from which source). When analyzing the *integration of semantic information* from sources in the text products three dimensions have been considered: 1) The way in which selected *semantic information evolved along the drafts*; 2) In which *succession* students arrange the *semantic information selected from sources*; 3) The *degree of integration* of semantic information selected from the given sources in the final products. The number of propositions was chosen as an indicator when quantifying the semantic information. A proposition was considered the equivalent of a proposition that caries semantic meaning independently, or of a construction of two propositions that need each others' completion in order to carry semantic meaning. For analyzing the *degree of integration of semantic information selected from the given sources* in the final products, the number of switches made between propositions selected from one particular source to propositions selected from another given source was used as indicator. The following types of switches were examined: between segments of propositions selected from the given sources (Sbb); between units selected from the same given source (Sb1b1); between a unit selected from a given source and a unit selected from another given source (Sb1b2). For examining the *degree of integration* between propositions selected from the given sources and propositions selected from other sources (others text books, or written sources found by students themselves, internet sources), the following switches were examined: between propositions selected from a given source and propositions selected from another source, no matter in which direction (SB0); between propositions selected from a given source and propositions selected from another source (Sb0); between propositions selected from another source and propositions selected from given sources (S0b); between propositions selected exclusively from other sources (S00).

The second type of analysis was *the assessment of quality of group products* and was conducted by using an assessment instrument. This instrument focused on three major aspects of the group products: use of sources, content and/of argumentation, and structure and language use. Every assessed aspect was rated on a three point scale (from 0 to 2). *Use of sources,* consisted of two elements: complete use of sources and copying and pasting. *Content and argumentation* considered what students actually did with the information from the given sources (or from other sources) in order to compose a coherent text. The component *Structure and language use*, focused on the organization of the text and on language use. An 80% total agreement between the two evaluators was reached when using this instrument, with a value of .66 for the Cohens kappa coefficient.

In order to answer the second research question, an *analysis of chat discussion protocols* was conducted. Based on iterative readings of the chat protocol a coding system was developed, consisting of the following coding categories: 1) Selection - statements about the selection of semantic information from the given sources (e.g., "..and for con-arguments that witches are blamed for everything"); 2) Integration - statements regarding the structure of the group products and the use of the selected semantic information in the texts (e.g. "…maybe we should say that witch prosecution was good, because if a witch is prosecuted other witches get to scared to go on with witchcraft…"); 3) Planning - statements regarding the planning of selection, integration and writing acts (e.g. " who is going to do the comparison?"); 4) Monitoring - statements regarding the monitoring of selection, integration and writing acts (e.g. " I begin with the arguments for assignment 3"); 5) Evaluation - statements regarding the evaluation of selection, integration, and the progress of the text (e.g. " no, there are too many arguments for witch prosecution"); 6) General Information - general statements on the progress of the group product and the writing process (e.g. " I'll look at some arguments from supporters, will you look at some from the opponents?"); 7) Other - any other statement not included in the other categories (e.g. " I will go on with chapter 3"). To select and code the text material we used the *Multiple Episode Protocol Analysis* (MEPA), a software program (Erkens, 2003) which offers the possibility to visualize and categorize the VCRI files.

Findings on Use of Semantic Information

Regarding the selection of semantic information we report on a number of aspects. Concerning the *quantity of semantic information selected from sources,* percentages of the propositions selected from sources show how quantity of information is distributed per draft, per group along the drafts and in the final product. The average quantity of semantic information selected from sources represents more than a third (35.9%), from the semantic information identified in the final products. A noticeable aspect is the high range between different groups. In both drafts and final texts the differences between groups in the selected number of propositions were big: in draft 1 the range is 79%; in draft 2 it is 65.9% and in the final text it is 58.5%. These differences show that some groups relied almost completely on the given sources for collecting semantic information for their assignment. Other groups, such as group 201 (9.9%), used especially information selected from other than the given sources.

Regarding the *type of semantic information from sources* on average 35.9% of the semantic content in the final group products is represented by information selected from the given sources. More than half (55.5%) of the selected and used propositions was core information; about a third (33.5%) was secondary information. These results show that groups concentrated on selecting *core information* and that students were less interested in selecting semantic information unrelated to the assignment. The *quantity of semantic information selected from each source* was quantified in relation to the total number of propositions per group product. The pie chart in Figure 2 shows that almost two thirds (63.8%) of the quantity of semantic information used for writing the final group products was selected from other sources than the six given sources.

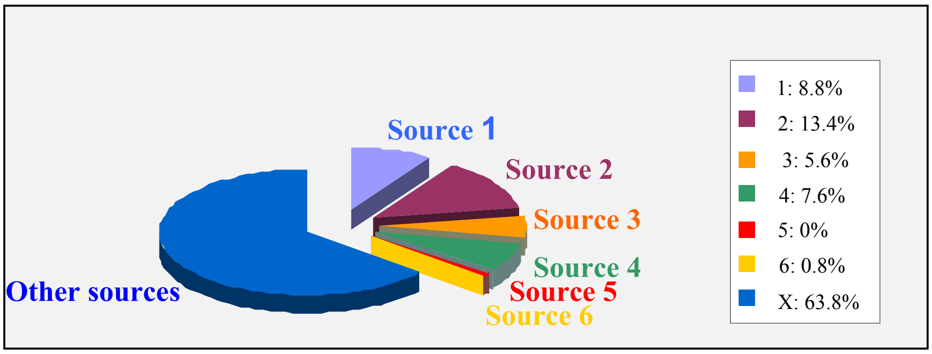


Figure 2. Quantity of Semantic Information Selected from Each Source and from Other Sources.

Only 35.2% of the propositions are selected from the given sources and the chart shows the division of the quantity selected from each source, in relation to the total number of propositions identified in the final group products. These proportions also indicate that groups had a strong preference for certain sources; especially for source 2 (13.4% from the total number of selected propositions). The content of the sources explains this preference: this source contains a great number of explicit pro- and con-arguments that could almost literally be used when writing the assignment text. Source 1 (8.8%), source 3 (5.6%) and source 4 (7.6%) contain a number of clear arguments as well. From these sources content was selected, but in a lower percentages. No units were selected from source 5. The percentages of propositions selected from other sources or invented by groups themselves are unexpectedly high in comparison with the percentages of units selected from the given sources.

When examining how students *integrated semantic information* selected from the given sources in their texts, a number of aspects were considered. The average quantity of semantic information selected from sources represents more than a third (35,9%) from the semantic information identified in the final group products; in the second draft, the percentage was four percent higher than in the final version (39.4%); and in the first draft half (50.3%) of the semantic content was selected from the given sources. It is noticeable that group 201 is the only group for which the percentage of the propositions selected from sources is higher in the final group product than the proportion in the first and second draft. For all the other groups this proportion is higher in the first draft and gradually becomes smaller, while the writing activities process heads towards the end. Regarding the *succession of the propositions selected from sources in the new text*, no stable succession pattern could be determined when looking at the organization of the units in each group product and when comparing the succession of the sources between the group products. Even so, some features of the succession could be named: six out of the ten groups started their text with propositions selected from source 2; in eight out the ten products source 2 was used in the first half of the texts; with one exception, sources 1 and 4 were used in the second half of the text; information units from source 3 and source 6 were randomly distributed in the semantic structure of the texts; information units from sources 1, 2, 3, and 4 were used on various locations in the texts.

To determine the *degree of integration,* two aspects were examined: the integration of only propositions selected from given sources and the integration of all propositions (from given and other sources). From the switches made between propositions selected exclusively from the given sources an average of 87.7 % were made between propositions selected from the same source; 12.3% were made between propositions selected from different given sources. This last percentage, together with the percentage indicating the integration of the segments of units, shows that the degree of integration of information units selected from different given sources was rather poor. Figure 3 presents the degree of integration per group in the form of switches between: units selected from given sources (Sbb, blue bar); units selected from given sources and units from non-given sources (SB0, purple bar); units from other sources (S00, white bar). With only one exception (group 203), the percentages of switches between units selected from given sources and the units from other (SB0) is higher than the percentage of switches between units from other sources (S00). Also, in five of the ten texts the percentage of switches between units from given sources and units from other sources (SB0) is lower than the units only selected from given sources (Sbb).

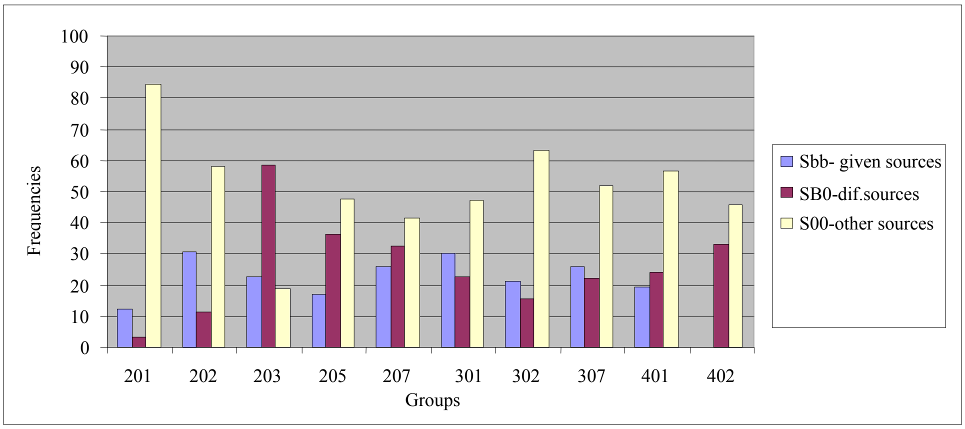


Figure 3. Percentage of Switches per Group.

Overall, the results show that 25.6% of the switches were made between units from different sources. Furthermore, 22.9% of the switches were made between units from the same source. Half of the total switches were made between units selected from other sources, and in this case there is no specification whether these propositions were selected from. The low percentage of switches between units selected from sources and units from other sources and the high percentage of switches between other sources supports the statement that the degree of integration is moderate.

A correlation analysis conducted to determine *the relation between quality of group products and the quantity and type of semantic information* shows a positive, but low ($r = .24$, $p < .10$) and non significant correlation between the obtained scores and the quantity of semantic information selected from sources. This is a correlation in the expected direction but does not confirm the expectation formulated at the beginning of this study, that groups which will select more semantic information from the given sources will also score better by the assessment of their final products. The result reveals that the obtained scores were influenced by the quantity of semantic information selected from sources, but not in a decisive manner. On the other hand, the correlation analysis *between the assessment scores and the proportion of core information selected from sources* showed a much higher degree of correlation. A positive correlation of .69 was found, and this correlation was also statistically significant ($p < .05$). The expectation that the groups that selected more semantic information labelled as core information will receive higher scores is confirmed by this result.

Findings on Meta-Cognitive Activities and Assessment of Group Products

The percentages calculated in relation to the total number of chat-statements indicate that statements about *Planning* occurred least frequently (0.4 %) and the statement on *Other aspects* most frequently (24.1%). The *Selecting* and *Integrating activities* (defining for the discourse synthesis activities) are discussed in 1.1%, respectively in 2% of the total statements. Also the discussion of *Evaluation* and *Monitoring* activities took place infrequently. In relation to the coded statements, the results show that more than half of the statements (58%) belonged to the category *Other aspects*. Statements coded as *General information* represented a third of the total. Planning, monitoring and evaluating were discussed in 9%, while selecting and integrating activities were only discussed in 7.6% of the coded statements. Furthermore, the statements regarding coded activities represented almost half percent (43.7%) of the total number of statements.

These results indicate that little discussion was held between the group members during the writing of the assignment. Observations indicated that group members split tasks and performed them without much consultation. The results show, in general, that the category *Other aspects* is strongly represented in the discussions of all groups. The discussion of one group (307) represents an exception from these findings. This group discussed about selection and integration of information from sources in their text. A qualitative analysis of this group's discussion showed that group members selected sequences of semantic information from sources individually and then discussed it with the other members, permanently discussed about the integration of selected information in the new text, and that the content, order of ideas, structure of the text, and conclusion were decided in consultation. Monitoring activities took place and, close to the end of the process, every group member performed a last evaluation of the text and discussed this with the group.

The relationship between *assessment scores* and *chat discussion categories* is reflected by a series of correlations. The correlations between the results and the categories *Selecting, Integrating, Evaluating, General information* and *Other* are positive, but low. The correlations between results and the categories planning and monitoring are negative. None of the correlations are statistically significant.

Table 1: Correlations between chat categories and a) assessment score and b) degree of integration.

Category	Correlations	
	Assessment scores	Degree of integration
Selecting	.28	-.20
Integrating	.32	-.19
Planning	-.36	-.41
Monitoring	-.13	.19
Evaluating	.13	-.12
General info	.22	-.04
Other	.26	-.31

These correlations show that discussing the selection and integration of information (from given or not-given sources) guaranteed a certain quality of the written text, and that also the fact that these activities took place is reflected in the better results obtained for assignment 3. As the results presented in Table 1 show, the correlations between the degree of integration of semantic information selected from sources and the meta-cognitive activities are, with one exception, negative. None of the correlations are statistically significant. These negative correlations indicate that there was no relationship between the frequency with which groups discussed about discourse synthesis activities and the degree of integration of propositions.

Summary of Findings, Discussion and Conclusion

In the present study we investigated the way groups of students use semantic information from multiple textual sources for writing their own text in collaboration. The first research question focused on how groups of secondary school students select semantic information from multiple sources when performing an inquiry task for the discipline history. When analyzing the quantity of semantic information selected from sources, the conclusion is that the differences in the quantity of selected information between groups were considerable. When distinguishing between the categories of semantic information offered by the given sources (core, secondary and other information), a selecting pattern was identified. All groups selected a higher quantity of core information, merely consisting of arguments pro- and counter witch prosecution. This category of semantic information was important for the quality of the final group products, as shown by the positive correlation between the assessment results and the categories of selected information. A much lower quantity of secondary semantic information was selected and used in the new text, and hardly any from the category of *Other information*. An aspect that deserves attention is the work strategy the majority of the groups used when performing this writing task. Firstly, semantic information from the given sources was selected. This information was organized based on contextual criteria (superordinate semantic categories), and not based on the source succession. The propositions selected from sources were organized in semantic sequences, most of the time units from the same source together. Thereafter, more sources of information were sought and new semantic information was selected from these sources. This information was also organized in semantic sequences, but it occurred more often that semantic information from these sources was intertwined with semantic information selected from the given sources. This semantic structure was completed by groups' own constructed propositions, which fulfilled the role of connecting propositions between semantic sequences. Regarding the degree to which students integrated semantic information from sources, the results showed that the degree of integration of semantic information from the given sources was situated at a quarter of the maximum degree possible. The result of such a structuring strategy is a lower degree of integration. The expectation that groups which use more information from sources will write better texts was not confirmed by the findings. The expectation that the groups that selected more core information will receive higher scores was confirmed.

The second research question concerned *meta-cognitive* activities initiated by students in order to regulate their information selecting and integrating process. The assumption was that statements concerning planning, monitoring, and evaluating the writing activities would be identified in the discussions students held through the chat. The results showed little presence of discussions regarding selecting and integrating semantic information from the sources. Additionally, other meta-cognitive activities which are characteristic for writing tasks in general, such as planning, monitoring or evaluating occurred rarely in groups' discussions. The content of the virtual discussions was dominated by statements regarding general aspects in relation with the task content and performance. The reason for this moderate representation of the meta-cognitive activities can be found in the fact that most of the groups chose to distribute tasks amongst the members of the group, each of them being responsible for one or more assignments. The expectation that there will be a relationship between the type (on- or off-task) and frequency of chat talk with the degree of integration in the final text product was not confirmed.

The characteristic of selecting information from sources based on the type of information these sources offer can be discussed in relation to the way students integrated the selected semantic information in the new text. The overviews of the semantic structure of the new texts show that groups selected semantic information from sources and organized it according to a content-based structure. According to the classification of rhetorical transforming made by Segev-Miller (1997), the groups synthesized the information selected from sources based on ideas rather than based on authors (sources). The strategy adopted by these groups can be classified as *decomposing and recomposing* the source texts, of breaking the source into propositions and looking across other text(s) for related propositions, which can also be placed under the same superordinate semantic categories.

A number of issues need to be considered in relation to this study. First, performing a task in a group context had certain consequences for the way a student approached the task and the responsibilities to be taken when performing a writing task. Moreover, when analyzing the products of a collaborative writing task the risk appears that the analysis of discourse synthesis activities performed at individual level would not be detailed enough. In this research project, the contribution of the individual participants is studied in the context of collaborative group work, which might overshadow some of the details of the individual input. On the other hand, the analysis of the writing activities when performed in a group context can offer extra insights into the mechanisms of the process. The necessity of consultation and agreement between the group members results in explication of individual ideas and stances, which provide valuable information about the collaborative process. Alas, in this research this potential advantage could not be used because of the reduced amount of explicit discussion on discourse synthesis activities. Second, the task used in this study was an authentic task used in the Dutch history curriculum, and not created especially for this investigation on discourse synthesis activities. Therefore, the sources were not selected with an eye on the discourse synthesis task. The texts were very complex and presented a high degree of difficulty. Especially the selection of information might have been affected by this complexity. It is possible that students avoided selecting information from the difficult sources and relied more on the rather accessible source texts. Third, an inter-rater reliability analysis should have been applied for establishing the reliability of the coding system for the chat protocols. Due to practical reasons this analysis was replaced by the researcher scoring the analyzed dialogues twice.

In sum, this study provided some valuable information on the way groups of students select and integrate semantic information from multiple textual sources, and the discussions of the groups while performing this task. Further research is needed for a more specific investigation of the degree of integration of semantic information from various sources, wherein the use of information from other than the given sources or students' prior knowledge are also considered. Moreover, further investigation of the group discussions around the task-related activities is recommended.

References

Erkens, G., Janssen, J., Broeken, M., Jaspers, J. (2002) *VCRI.* Retrieved March 23 from Utrecht University, department of Educational Sciences Web site: http://edugate.fss.uu.nl/~crocicl/vcri.html.

Erkens, G. (2003). Multiple episode protocol analysis (MEPA).Version 4.9. Utrecht University, The Netherlands.

Higgins, L. (1992). Reading to argue: helping students transform source texts. In A. Penrose & B. Sitko (eds.), *Hearing ourselves think: cognitive research in the college writing classrooms.* (pp. 77-101). New York: Oxford University Press.

Kantz, M. (1989). *Written rhetorical synthesis: processes and products*. (Technical report no. 1). The Center for the Study of Writing, Pittsburgh, PA & Berkeley, CA, Retrieved on August 15, 2005, from National Writing Project, U.S. Department of Education web site: http://www.writingproject.org/TR17.pdf.

Kintsch, W. &Van Dijk, T. (1978). Towards a model of text comprehension and production. *Psychological Review*, *85*, 363-394.

McGinley, W.J. (1992). The role of reading and writing while composing from sources. *Reading Research Quarterly, 27 (3),* 226-249.

Segev-Miller, R. (in press). *Cognitive processes in discourse synthesis. On the differences between successful and unsuccessful synthesizers*. A paper presented at the 9th International Conference on the EARLI on Writing, Geneva, Zwitserland (September, 2002).

Segev-Miller, R. (1997). Cognitive processes in discourse synthesis: The case of intertextual processing strategies. In *Cognitive processes in discourse synthesis*. (Dissertation). The Hebreuw University of Jerusalem, Retrieved on Mei 15 2005 from http://www.staffs.ac.uk/personal/sciences/mt15/writing.doc.

Segev-Miller (2004). Writing from sources: the effect of explicit instruction on college students' processes and products, in *Educational Studies in language and literature*, *4*, 5-33.

Spivey, N.N., & King, J.R. (1989). Readers and writers composing from sources. M.R. Ruddell (ed.). In *Theoretical models and processes of reading*, (pp. 668-694). Newark: International Reading Association.

Spivey, N.N. (1995). Written discourse: a constructivist perspective. In Steffe, L.P. & Gale, J. (1995). *Constructivism in education*. Hillsdale: Lawrence Erlbaum Associates.

Spivey, N.N. (1997). *The constructivist metaphor. Reading, writing and the making of meaning*. San Diego: Academic Press.

Online Role Play Simulation to Tackle Groupthink – Case Study of a Crisis Management Training

Lawrence Leung, Nancy Law, The University of Hong Kong, Pokfulam Hong Kong
Email: h8690182@hku.hk, nlaw@hku.hk

Abstract: Decision-makers have to face many challenges during crisis management: ill-structured problem, time stress, incomplete information and involvement of multiple parties. Groupthink is a term coined by psychologist (Janis, 1972), Groupthink victims ignore alternatives and tend to make irrational decisions in facing crises. Many researchers suggest that role play simulation (RPS) is a suitable tool for crisis management training for the realisation of collaboration and authentic situation awareness. In this paper, a case study of crisis management training in the context of law enforcement using online RPS is illustrated. By using Bales' Interaction Process Analysis (IPA) method, the interactions of participants in training are observed and categorised. The results show that using online role play simulation in crisis management training has positive effect on the reduction of Groupthink tendency.

Introduction

Crisis management is a set of actions taken to exert control over the events of a crisis to minimize losses. It is similar to risk management, except that the events are real, not potential, and action takes precedence over planning (Sniezek, Wilkins & Wadlington, 2001). Crisis management, to be effective, involves making appropriate decisions under severe time pressure and uncertainty. The objective of crisis management training is to enhance the ability to solve problems by adapting very quickly to a fast changing situation. Training in crisis management is essential to many organisations nowadays, especially law enforcement agencies. Crisis management is a highly complex subject and decision-making is a core process of it. Decision-making is critical in nature and also ill-structured, involving multiple parties, a lot of information from various sources. Moreover, the decision has a very short time frame to make, but the decision may cause a devastating consequence. One important challenge in crisis management training is inducing the psychological processes that associated with the acute stress experienced in real-world crises. Acute stress is a state that occurs in situations of potential harm, time pressure, and arousal (Sniezek, Wilkins & Wadlington, 2001). A most important feature of decision-making in crisis management is that it is group activity, dynamic in nature and hard to teach.

Crisis decision-making in law enforcement context is considered to be naturalistic. Naturalistic Decision Making (NDM) research emerged in the 1980s by Gary Klein to study how people make decisions in real-world settings. Realism is important, as lack of realism can prevent transfer of training, especially for performance under time pressure (Zakay & Wooler, 1984). The NDM framework emphasizes the role of experience in enabling people to rapidly categorize situations to make effective decisions (Klein, Orasanu, Calderwood, & Zsambok, 1993). Many researchers agree that role play simulation is a suitable method for crisis management training (Klieboer, 1997; Friedman 2004; Borodzicz & Haperen, 2002; Borodzicz, 2005; Miller, 2009). Their arguments include: offering trainees a setting approximating real-life experiences to learn how to apply insights from crisis management theories, supporting "learning by doing" in a safe environment, engaging collaboration and adopting scenario-based training. Effective crisis management simulations encourage participants to perceive the scenario as a threat, with time limitations for effective data gathering. Simulations should produce the similar reactions and feelings in participants as real life crisis events, such as tension, uncertainty, time pressure, sense of inadequate information and frustration (Gredler, 1992).

Groupthink

Groupthink was originated by Irving Janis in his book Victims of Groupthink in 1972, and explained how group made poor decisions during the decision-making process. Janis coined the term "Groupthink" to refer to "a mode of thinking that people engaged in when they are deeply involved in a cohesive in-group, when members' striving for unanimity override their motivation to realistically appraise alternative courses of action." In summary, Groupthink occurs when the pressure to conform within a group interferes with that group's analysis of a problem and causes poor group decision-making. Groupthink victim's creativity, uniqueness, and independent thinking are lost in the pursuit of group cohesiveness. The hypothesis was supported by Janis's hindsight analysis of several political-military fiascos and successes that were differentiated by the occurrence or non-occurrence of antecedent conditions, Groupthink symptoms, and decision-making defects. The political-military fiascos that Janis studied including: Nazi Germany's decision to invade the Soviet Union in 1941, the Pearl Harbour Incident, the Bay of Pigs Crisis, North Korean Invasion, the escalation of the Vietnam War and the Watergate Cover-up.

In groupthink research, case study, laboratory experiment and content analysis are the major methods used. However, it can be argued that online role play simulation should be more comprehensive in facilitating decision-making training in crisis management or avoidance of groupthink in specific. The arguments include: (i) Decision-making training concentrates on process not outcome alone and (ii) Decision-making is not a singularity issue, but combination of multiple decision types and events (Orasanu & Fischer, 1997). In fact, these arguments align with Mohamed & Wiebe (1996)'s conclusion in Groupthink research that, Groupthink is a continuous process, rather than a list of discrete events. Furthermore, the process flow of Groupthink is non-deterministic and probabilistic instead of deterministic or statically formulated. In other words, it is not necessary for the entire antecedent conditions have to be present for Groupthink to occur. Rather, the more antecedent conditions that are present the greater the risk of developing the symptoms of Groupthink (Rosander, Stiwne and Granstorm, 1998). Moreover, Esser and Lindoerfer (1989) have revealed the similar findings with Rosander, Stiwne and Granstorm. Therefore, it can be concluded that more Groupthink symptoms then lead to a greater probability of defective decision-making.

Tackling Groupthink with Online Role Play Simulation

Role Play Simulation (RPS) is being increasingly appreciated as a useful technique to foster virtual learning processes (Lombard & Biglan, 2009). RPS takes inspiration from "situated constructivism" approaches, which state that an educational experience has to be as authentic and genuine as possible, so that learners can observe and critically reflect on real situations (Winn, 1993). Moreover, RPS is directly rooting from the situated learning approaches of Goal-Based Scenario (Schank, 1997); where the learner is engaged in pursuing a goal within a simulated environment in order to master a set of skills. Role Plays are techniques based on such principles; which are increasingly being adopted in CSCL contexts (Persico D., Pozzi F. & Sarti L., 2008). In fact, RPS is a new area in CSCL research that professional learning can apply.

Online role play simulation, a special form of CSCL, is complied of a set of technologies in an online environment for collaborative learning. Online RPS is designed to support and structure group interactions for the purpose of information exchange, problem solving and decision-making. Through the development of various scenarios, the crisis decision process can be realised and monitored. Online role play simulation supports activities such as idea creation, message exchange, project planning, document preparation, and joint planning and decision-making that Group Support System (GSS) usually provides (Poole & DeSanctis, 1989). Moreover, features of online role play simulation are believed to break down hierarchies that dominate typical meetings (Zigura, Poole & DeSanctis, 1988) and to equalise the participation of all group members. In terms of the context of group decision support, online role play simulation usually consists of networked computers with monitors for group members in simulating crisis decision-making and aggregating group opinions. Researches have proved that computer technology based system can help mitigate the negative effects of antecedent conditions of Groupthink. These systems usually include process structuring, a public screen, anonymity, simultaneity, extended information processing, and access to external information (Miranda, 1994; Chidambaram, Bostrom & Wynne, 1991). System with anonymity, simultaneity, process structuring, and the public screen can foster productive conflict management (Miranda, 1994). As those features contribute to a task-focus thereby preventing unproductive conflict that is unrelated to the issue at hand. Those features also promote group members' freedom to participate in the discussion. As such, a high amount of issue-based conflict results. As a matter of fact, such constructive conflict can also promote an understanding of the problem and of opposing frames of reference as well as an exploration of alternatives. Consequently, a greater differentiation among group members' opinions may occur. When a group possesses Groupthink antecedents, it may not be possible or even desirable, to change the group's structure. However, it is possible to prevent the development of Groupthink by ensuring that each specific group decision is made under conditions of vigilance (Janis, 1982). This implies that the group examines a wide range of alternatives, carefully evaluates and re-evaluates these alternatives within the context of the group's objectives, search for relevant information, accurately processes all information, and undertakes detailed implementation and contingency planning.

Using Interaction Process Analysis to Understand Groupthink Processes

The Bales' (1950) IPA method and theory work on a relationship between social-emotional and tasks-based communication acts. It is generally agreed by researchers that there are two main dimensions of group life, task and social aspects (Frey, 1999), and these two dimensions are used to accomplish tasks and maintain group relationships. Bales' IPA can provide a detailed observational scheme at micro level for coding group members' communicative behaviour, such that it can be recorded, isolated and interpreted. By coding member's behaviour into discrete categories, researchers are able to interpret whether participants' comments are helpful or disruptive to the group and whether the communication acts are balanced (Schultz, 1999). Other coding schemes explored, Roter's (1991) (doctor-patient) and Eyberg & Robinson's (1983) (parent-child), are applied to specific domain of group interactions and are not appropriate for this study. In Bales's IPA, spoken or written communication will be broken into "units" consisting of a single thought or the equivalent of a simple sentence.

According to coding scheme, each communication or interaction unit can be assigned to one of 12 mutually exclusive categories. Details of the Bales' IPA coding scheme are illustrated in Table 1.

The Watergate cover up was confirmed as a Groupthink case by Cline (1994). Cline used Bales' IPA to analyse transcripts of the Watergate cover up case that were known to have made flawed decisions and found that too much emphasis on agreement resulted in unsuccessful outcomes. In "real life", groups that were found to be associated with Groupthink, the level of agreements were ten times higher than disagreement and, when Groupthink occurred in laboratory studies, agreement was seven times greater. Furthermore, Cline highlighted three components on contributing/avoiding Groupthink – Agreement, Task-based and Disagreement. According to Cline's research, too much agreement might decrease the chance of constructive conflict in decision-making; which was considered to be healthy. It was because under conflict situation more alternatives could be reviewed and decisions could be fully reviewed. Cline also found out that decision-making process was more task-based than socio-emotional based, as the discussion process was mainly focus on the solution finding, information gathering and sharing.

Table 1: Bales' 12 interaction categories (adopted from Bales,1950, p.9).

Code	Category	Remark
1	**Shows solidarity** – Raises other's status, gives help, encourages others, reinforces (rewards) contribution, greets others in a friendly manner, uses positive social gesture.	Social Emotional Area: Positive Reactions
2	**Shows tension release** – Jokes, laughs, shows satisfaction, relives or attempts to remove tension, expresses enthusiasm, enjoyment, satisfaction.	
3	**Agrees** – Shows passive acceptance, acknowledges understanding, complies, co-operates with others, expresses interest and comprehension.	
4	**Gives suggestion** – Makes firm suggestion, provides direction or resolution, implying autonomy for others, attempts to control direction or decision	Task Area: Neutral
5	**Gives opinion –** Offers opinion, evaluation, analysis, express a feeling or wish. Seeks to analyses, explore, enquire. Provides insight and reasoning.	
6	**Gives information –** Provides background or further information, repeats, clarifies, confirms. Brings relevant matters into the forum, acts that assist group focus.	
7	**Asks for information –** Asks for further information, repetition or confirmation. Acts used to request relevant information and understand the topic.	
8	**Asks for information –** Asks for further information, repetition or confirmation. Acts used to request relevant information and understand the topic.	
9	**Asks for suggestion –** Asks for suggestion, direction, possible ways of action. Requests for firm contribution, solution or closure to problem.	
10	**Disagrees** – Shows passive rejection, formality, withholds help, does not support view or opinion, fails to concur with view, rejects a point, issue or suggestion.	Social Emotional Area: Negative Reactions
11	**Shows tension** – Shows concern, apprehension, dissatisfaction or frustration. Persons interacting are tense, on edge. Act that express sarcasm or are condemning.	
12	**Shows antagonism** – Acts used to deflate others status, defends or asserts self, purposely blocks another or makes a verbal attack.	

Previous studies that have successfully used the IPA method to investigate various aspects of communication including: Hiltz and Turoff (1993) - comparison between face-to-face and computer mediated communication (CMC), Cline (1994) - Watergate cover up case research, and Gorse and Emmitt (2003) - construction meeting using CMC.

Study Design

Research Questions

1. In what ways does socio-emotional and task-based communications affect Groupthink tendency?
2. Can participation in Online Role Play Simulations be an effective means to reduce Groupthink tendency?

Method

This study involved 16 trainees from a law enforcement training institute in Hong Kong using role play simulation exercise as part of the 4-week crisis management training programme. In this exercise, trainees were randomly assigned to 4 different teams namely: Alpha (α), Beta (β), Gamma (γ) and Delta (δ) with specific functional roles in a crisis scenario. Each team was located in different rooms of the exercise vicinity without face-to-face contact. The facilitator or exercise controller at the control room could communicate with the trainees of all groups through a computer networked role play simulation platform called SIMS (Scenario-based Interactive Role Play Simulation System). Participants did not know the scenario of the role play beforehand until the commencement of the exercise. SIMS provides all the simulated communication tools for law enforcement training, including chat-box, e-mail, beat radio, telephone and video conferencing system. Trainees can use the tools provided to carry out decision-making and Command & Control (C&C) operation during the role play simulation. In addition, Controller can also send out instant multimedia information (such as TV news, video clips) according to scenario development or enhancing realism of the crisis situation. All role play interactions and communications are logged by the system for data analysis and debriefing.

Procedure

The whole exercise was divided into two phases with two debriefings. Both Phases I and II were role play simulation, each phase had the same duration of 2 hours. At the beginning of the exercise, participants had to complete the 40-question Groupthink Index (Glaser, 1993) questionnaires, which were used to measure their Groupthink tendency before the exercise. Debriefing I was conducted after the completion of Phase I, in this 30 minutes session, the facilitator delivered a short introduction about Groupthink to participants. After a break, Phase II commenced. The Debriefing II was the post-exercise discussion. Facilitator discussed with participants about the exercise that was conducted. After the exercise, participants completed the Groupthink Index questionnaires again to understand the change of Groupthink tendency.

Data Collection and Method of Analysis

The interactions of the exercise were recorded by SIMS, including text (chat-box, e-mail), audio (simulated beat radio, telephone) and video (video conference). All the recorded transactions were then transcribed for analysis. Transcripts were coded according to IPA and in order to examine the impact of status differentials, a percentage (%) figure indicating the relative number of "units" by each team was computed to give an indication of their proportionate contribution to the overall interactions. Representative sections (10%) of text were double-coded by two independent raters to assess inter-rater reliability. The guidelines for IPA coding were clear and consequently agreement between raters approached 97%.

Result and Analysis

Measurement of Groupthink Tendency

Glaser's Groupthink Index (1993) is the instrument used in this paper. According to Glaser, the significance of Groupthink in decision-making can be divided into different categories by severity, it is suggested that an overall score (over the eight 'Groupthink symptoms') of less than 93 indicates 'Very insignificant' groupthink tendency; 94-111 indicates 'Insignificant' groupthink tendency; 112-129 indicates 'Moderate' groupthink tendency; 130-147 indicates 'Significant' groupthink tendency; and a score greater than 148 represents 'Very significant' groupthink tendency.

The change of Groupthink Index of each team, between the Phase I score and the Phase II score were: Team Alpha (-15.11%, Groupthink tendency reduced from High to Moderate), Team Beta (-8.48%, Groupthink tendency reduced from High to Moderate), Team Gamma (-17.88%, Groupthink tendency reduced from High to Moderate) and Team Delta (-20.49%, Groupthink tendency reduced from High to Moderate). As a whole, all teams had decrease in Groupthink tendency from High to Moderate but in different degrees. Interestingly, the performance of Team Beta was significantly lower than the other 3 teams. The reduction of Groupthink tendency indicates that online RPS may be a factor to trigger the change. However, it is not sufficient to reach a conclusion at this point, unless the positive effect between online RPS and change of Groupthink tendency is established. Therefore, the following IPA analysis is to investigate this relation. The coded profile of the interactions are summarised in Table 2 in sorting order of mean. From Table 2, it can be observed that the IPA category with highest mean is "Gives information (6)" and the lowest mean is "Shows antagonism (12)". The table reflects that exchange of information, suggestion and opinion are the dominant activities, compare to other social-emotional interactions. In fact, this implies that crisis decision-making has gone through the process of information gathering, discussion and team work. This distribution of IPA codes is significantly different from the study by Bales (1950), which was a face-to-face free form play. However, the distribution of IPA codes is similar to the study by Hiltz (1978) in certain extent on problem solving via computer conference. To address the research questions, two levels of analysis are performed. The first level analysis is on the relationship between socio-emotional/task-based communications. The second level analysis is on the relationship between the 4 interaction categories (Positive/Negative Socio-emotional and Giving/Asking Task-based) and the Groupthink tendency.

Table 2: Bales' IPA categories – sorted by mean.

IPA Category (code)	Mean	Std. Dev.	Team (%) – in % order
Gives information (6)	45	17.166	γ (26.55), α (20.29), δ (17.35), β (15.17)
Asks for information (7)	37.25	7.5	β (18.62), δ (17.51), γ (17.26), α (16.30)
Gives suggestion (4)	25	10.296	α (13.77), δ (11.98), γ (10.18), β (8.97)
Gives opinion (5)	24.75	10.079	δ (12.8), α (12.68), γ (11.06), β (7.59)
Asks for opinion (8)	24.25	3.304	β (15.17), γ (12.39), δ (9.68), α (9.42)
Asks for suggestion (9)	22.5	2.517	β (15.86), γ (11.06), δ (8.76), α (8.33)
Agrees (3)	18	5.354	δ (10.6), β (8.33), α (7.97), γ (6.64)
Disagree (10)	8.5	5.916	δ (15.5), α (5.4), β (2.07), γ (1.77)
Shows solidarity (1)	4.5	2.646	β (5.56), α (1.81), γ (1.33), δ (0.92)
Shows tension release (2)	3.75	2.217	α (2.54), β (2.07), δ (1.38), γ (0.88)
Shows tension (11)	2.25	1.708	α (1.45), δ (1.38), γ (1.98), β (0)
Shows antagonism (12)	0.25	0.5	β (0.69), α (0), γ (0),, δ (0)

First Level Analysis - Social-emotional and Task-based Interactions

Interaction data gathered using the Bales' IPA can be studied from a number of different perspectives. In this paper, the first level of analysis is the social emotional and task-based components of communication and their relationship with the change of groupthink tendency. There are 6 social-emotional acts (that is shows solidarity, shows tension release, agrees disagrees, shows tension and shows antagonism) and the 6 task-based communication acts (that is gives suggestion, gives opinion, gives information, asks for information, asks for opinion, and asks for suggestion).

Figure 1 presents the interaction for all teams observed during the role play simulation exercise. It shows consistent patterns of interactions during the exercise. The percentage between Socio-emotional and Task-based interaction are 17.25% and 82.75% respectively. The distribution of categories shows the pseudo-normal pattern typically found in task-oriented group (McGrath, 1984). Categories relating to task performance (4-9) occur more frequently than those relating to socio-emotional behaviours (1-3) and (10-12). Interactions concerned with giving information (4-6) are more frequent than that aim at eliciting information (7-9). It can be expected that task solution involves information exchange and interpretation and where roles during decision-making processes largely involve reporting of information gathered by individuals between teams. Positive socio-emotional interactions (1-3) are more common than negatives categories (10-12), suggesting a healthy level of interpersonal interaction. This distribution can reveal that crisis management interaction is more task-based, with 43.92% on Giving tasks (Categories 4,5&6) and 38.94% on Asking tasks (Categories 7,8&9). On the other hand, socio-emotional tasks contribute 12.17% on Positive socio-emotional (Categories 1,2&3) and 4.89% on Negative socio-emotional (Categories 10,11&12). Of all the four teams, Team Beta has the most interactions on Shows solidarity (Category 1) with 5.52% and the least interaction percentage on Disagree (Category 10) (2.07%) and Shows tension (Category 11) (0%). In addition, Team Beta has also the least percentage on Task-based interaction (13.33%) among all four teams. According to Cline (1994), Team Beta has shown a high degree of socio-cohesion, which can contribute Groupthink among team members and they do not engage in conflict during the exercise. However, conflict in discussion is considered to be useful for Groupthink avoidance (Cline, 1994).

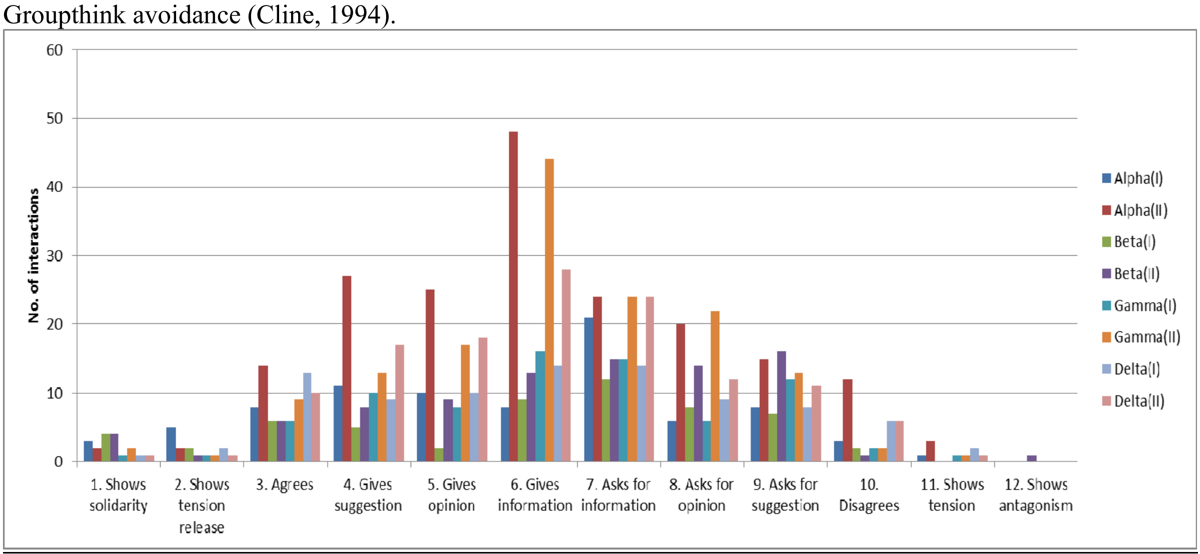


Figure 1. IPA Profiles of Teams.

To further understand the interaction pattern of Team Beta, the IPA profile in Phase I and Phase II are illustrated in Figure 2.

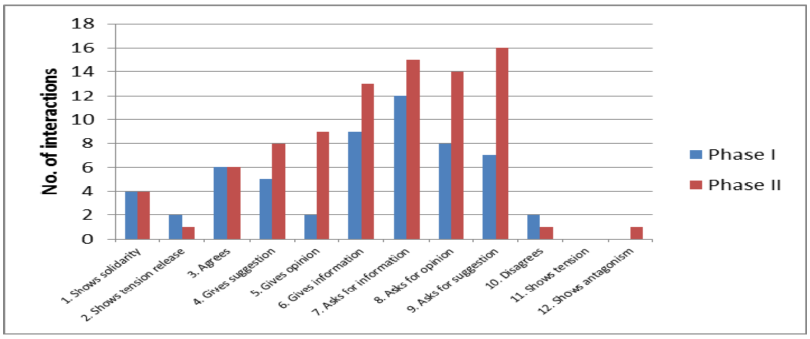


Figure 2. Team Beta (IPA Profile in Two Phases).

It can be revealed from Figure 2 that some decrease on Positive socio-emotional interaction – Shows tension (Category 2) and Negative socio-emotional interaction – Disagree (Category 10) and Shows antagonism (Category 12). However, not much improvement on Show solidarity (Category 1), Agrees (Category 3) and Shows tension (Category 11). It can also be seen that the major improvement is on task-based interaction, especially for the 4-9 categories.

Figure 3 is a graphical comparison of various Groupthink studies by Cline (1994) and the "Crisis Management" of this study. In this graph, the three major components on contributing/avoiding Groupthink are displayed in terms of percentage of interactions. It can be seen that "Crisis Management" has the least percentage (8.33%) of interaction on Agreement, which refers to Category 3 of IPA. According to Cline's research, too much agreement may decrease the chance of constructive conflict in decision-making; which is considered to be healthy. It is because under conflict situation more alternatives can be reviewed and decisions can be fully reviewed. In addition, the percentage of task-based interaction of Crisis Management (82.75%) is also the highest among other studies. It can observed that due to anonymity, simultaneity, extended information processing, and access to external information of the online role play simulation environment, participation of group members is encouraged. Such that the hierarchies or process structures that dominate traditional meetings can be broken. Furthermore, the percentage of Disagreement (Category 10) interactions is 3.94%, also lowest among the studies. "Crisis Management" scores a roughly 2:1 Agreement and Disagreement ratio, which implies a low Groupthink tendency according to Cline.

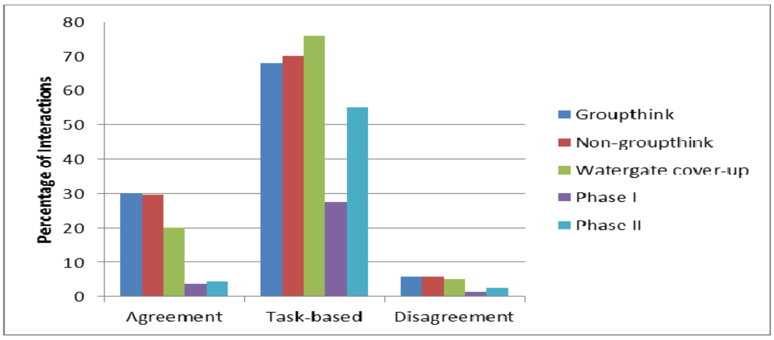


Figure 3. Comparison of IPA Profiles on Groupthink Studies.

Secondary Level Analysis – Interactions vs Change of Groupthink Tendency

The second level of analysis looks into the social-emotional and task-based components for more details. It is necessary to find out whether the 4 interaction categories (associated with the positive and negative social-emotional interaction, and giving and asking task-based interaction) are significantly related to the change of groupthink tendency. In this investigation, the interactions collected are divided into two groups, the successful group, that is transactions from teams α, γ and δ, and unsuccessful group, that is transactions from team β. To examine the social-emotional and task-based categories, individual IPA categories are grouped together accordingly. Positive social-emotional categories comprise of the aggregated data from Category 1 (Shows solidarity), 2 (Shows tension release) and 3 (Agrees). Negative social-emotional categories are also split into two groups – giving and asking task-based interaction. Giving task-based interaction included Category 6 (Gives information), 7 (Gives opinion) and 8 (Gives suggestion). Asking task-based interaction includes Category 7 (Asks for information), 8 (Asks for opinion) and 9 (Asks for suggestion). Table (3) provides a summarised comparison and statistical analysis of social-emotional and task-based interaction against the change of

Groupthink tendency. Finally, the Pearson's Chi square (χ^2) is used to examine the relationship between interaction levels and the change of Groupthink tendency. The findings are reported as probability values (ρ), the degrees of freedom are also stated. The findings are used as evidence supporting or rejecting of the hypothesis. The hypothesis is based on the acceptance that the level of interaction had a significant relation associated with the change of Groupthink tendency. In this paper, a significant difference is recorded as ρ =< 0.05. The result shows that χ^2 is 14.994, degree of freedom (df) = 3 and ρ value of < 0.02, which is considered to be highly significant. Therefore, the hypothesis is accepted, such that Groupthink tendency reduction is highly related with the Positive/Negative socio-emotional and Giving/Asking task-based interactions.

Table 3: Results of positive and negative socio-emotional interaction, and giving and asking task-based interaction: Groupthink tendency reduction successful and not successful. (SPSS output).

Groupthink tendency reduction	Pos. Social-emotional	Neg.Socio-emotional	Giving Task-based	Asking Task-based	Total
Successful	82	40	333	264	719
Not successful	24	4	46	71	145
Total	106	44	379	335	864

In order to demonstrate the relationship of the hypothesis with the full set of IPA categories, a further analysis is carried out, and the results are tabulated in Table 4.

Table 4: Results of 12 IPA categories: Groupthink tendency reduction successful and not successful. (SPSS output).

Groupthink tendency reduction	IPA1	IPA2	IPA3	IPA4	IPA5	IPA6	IPA7	IPA8	IPA9	IPA10	IPA11	IPA12	Total
Successful	10	12	60	87	88	158	122	75	67	31	9	0	719
Not successful	8	3	12	13	11	22	27	22	23	3	0	1	145

The result shows that χ 2 is 31.842, degree of freedom (df) = 11 and ρ value of 0.002 (Assmp. Sig (2-sided))< 0.05, which is also considered to be highly significant. Therefore, the hypothesis is accepted, such that Groupthink tendency reduction is highly related with the 12 IPA categories.

Conclusion

From the result of the analysis, it is shown that participants have achieved a reduction of Groupthink tendency after the Online RPS exercise. Moreover, the results of the IPA analysis have also proved that online RPS has positive effect in the reduction of Groupthink tendency. Online RPS has the potential to provide scenario planning that is a cost effective supplement to traditional face-to-face training. This new form of planning simulation provides flexibility and ease of preparation while stimulating creative and critical thinking. The design of SIMS aligns with the NDM framework, such that SIMS allows representation of actual crises and provides immersive interface contributes to acute stress. With the adoption of multimedia technology, the audio-visual effects and simulated tools of SIMS enhance the degree of realism. In addition, critiques and arguments are allowed; mode of interactions is multi-way and without hierarchy. Moreover, trainer can assess trainee actions instantly and intervene constructively. Trainer has made use of the logged interactions in SIMS to discuss and elaborate a specific concept or knowledge among trainees during debriefing, which cannot be achieved without technology. This paper describes the procedure to use Bales' IPA for analyzing electronic discussions, and the analysis demonstrates that the group is clearly task-based and discussions are mainly concerned with the exchange of information or opinion. With a closer inspection of data or discussion dialogues, the group is characteristically by generally positive interpersonal relations, with little antagonism or conflict evident. The group functioned well in terms of task content and process. However, it should be acknowledged that the data generated are limited in scope and a more detailed content analysis based on qualitatively derived coded may be necessary to fully capture the detailed characters of the group.

References

Bales, R. F. (1950). Interaction process analysis. Cambridge, MA: Addison-Wesley.

Borodzicz, E., & Haperen, K. V. (2002). Individual and group learning in crisis simulations. *Journal of Contingencies and Crisis Management,* 10(3), 139-147.

Borodzicz, E. P. (2005). Risk, Crisis and Security Management. West Sussex, England: John Wiley &Sons Ltd.

Chidambaram, L., Bostrom, R. P. & Wynne, B. E. (1991). A Longitudinal Study of the Impact of Group Decision Support Systems on Group Development. *Journal of Management Information Systems*, 7(3).

Cline, R. J. W. (1994). Groupthink and the Watergate cover up. The illusion of unanimity. In L. R. Frey (Ed.) *Group Communication in Context: Studies of Natural Groups*. New Jersey: Lawrence Erlbaum.

Esser, J. K., & Lindoerfer, J. L. (1989). Groupthink and the space shuttle Challenger accident: Toward a quantitative case analysis. *Journal of Behavioral Decision Making*, 2**,** 167–177.

Eyberg, S.M., & Robinson, E. (1983). Dyadic parent-child interaction coding system: A manual. Psychological Documents, 13, 2424, Ms. No. 2582.

Friedman, S. (2004), Learning to make more effective decisions: changing beliefs as a prelude to action, The Learning Organization, 11(2), 110 – 128.

Frey, L.R. (1999). The handbook of group communication theory and research. London. Sage.

Glaser, R. (1993). Groupthink Index, Organizational Design and Development, King of Prussia.

Gorse, C. & Emmitt (2003). Investigating interpersonal communication during construction progress meetings: Challenges and opportunities. Engineering,Construction and Architectural Management, 10(4), 234-44.

Gredler, M. E. (1992). Designing and evaluating games and simulations. London: Kogan Page.

Hiltz, S. R., Johnson, K., & Agle, G. (1978). Replicating Bales' problem solving experiments on a computerized conference: A pilot study (Research Report No. 8). Newark: New Jersey.

Hiltz, S. R. et Turoff, M. (1993). Social and Psychological Processes in Computerized Conferencing. In The Network Nation: Human Communication via Computer. MIT Press, Cambridge, MA.

Janis, I. L. (1972). Victims of Groupthink, Houghton-Mifflin, Boston.

Janis, I. L. (1982). Groupthink, 2nd ed. Houghton-Mifflin Company, Boston.

Klein, G. Orasanu, J. Calderwood, R. & Zsambok, C.E. (Eds.) (1993). Decision Making in Action: Models and Methods (pp. 327-345). Norwood, NJ: Ablex.

Kleiboer, M. (1997). Simulation methodology for crisis management support JCCM 5(4): 198-206

Lombard, R. and Biglan, B. (2009). Implications of Role Play and Team Teaching as Strategies for Information Technology Pedagogy, Information Systems Education Journal, 7(20).

Miller, L. (2009). Psychological Principles and Practices for Superior Law Enforcement Leadership in Police Chief.

Miranda, S M. (1994). Avoidance of groupthink. Small Group Research, 25(1), 105-136.

McGrath, J.E. (1984). Groups: Interaction and Performance, Englewood Cliffs, N.J: Prentice-Hall.

Mohamed & Wiebe (1996). Toward a process theory of Groupthink. *Small Group Research*, 27(3), 416-431.

Orasanu, J. & Fischer, U. (1997). Finding decisions in natural environments: the view from the cockpit. In C.E. Zsambok & G. Klein (Eds.), Naturalistic Decision Making . Mahwah, NJ: Lawrence Erlbaum.

Persico, D., Pozzi, F., & Sarti, L. (2008). Fostering collaboration in CSCL. In A. Cartelli, M. Palma (Eds.), Encyclopedia of information communication technology, (pp. 335-340). Idea Group Reference.

Poole, M.S., & DeSanctis, G. (1989). Use of Group Decision Support Systems as an Appropriation Process, Proceedings of the Twenty-Second Annual Hawaii International Conference*,* 4, 149-157.

Rosander, M., Stiwe, D., & Granström, K. (1998). Bipolar Groupthink: Assessing groupthink tendencies in authentic work groups. Scandinavian Journal of Psychology, 39, 81-92.

Roter, D.L. (1991). The Roter Method of Interaction Process Analysis (RIAS Manual).

Schank, R. (1997). Virtual Learning: A revolutionary approach to building a highly skilled workforce. New York: McGraw-Hill.

Schultz, B.C. (1999). Improving group communication performance, an overview of diagnosis and intervention. In*:* L.R. Frey. The handbook of group communication theory and research. London, Sage. 371-394.

Sniezek, J. A., Wilkins, D. C., and Wadlington, P. (2001). Advanced Training for Crisis Decision Making: Simulation, Critiquing, and Immersive Interfaces, *Proceedings of the 34th Hawaii International Conference on Systems Science*s, Maui, Hawaii. January 5-7, 2001.

Winn, W. (1993). A constructivist critique of the assumptions of instructional design. In T. M. Duffy, J. Lowyck, & D. H. Jonassen (Eds.), Designing environments for constructive learning, 189-212. Berlin: Springer-Verlag Conference on System Sciences (HICSS-34), 3, 2001.

Zakay, D., & Wooler, S. (1984). Time pressure, training and decision effectiveness. Ergonomics, 27 (3), 273 - 284.

Zigurs, I., Poole, M. S., & DeSanctis, G. (1988). A study of influence in computer mediated group decision making. MIS Quarterly, December, 625-644.

Play and Augmented Reality in Learning Physics: The SPASES Project

Noel Enyedy, UCLA, Moore Hall 2323, 405 Hilgard Ave, LA , CA 90095, enyedy@gseis.ucla.edu
Joshua A. Danish, Indiana University, 201 North Rose Ave, Bloomington, IN, 47405, jdanish@indiana.edu
Girlie Delacruz, UCLA, Moore Hall, 405 Hilgard Ave, LA, CA 90095, gdelacruz@cse.ucla.edu
Melissa Kumar, UCLA, Moore Hall, 405 Hilgard Ave, LA, CA 90095, mjdkumar@ucla.edu
Sylvia Gentile, UCLA Lab School, 330 Charles Young Dr N, LA, CA, 90095, sgentile@ucla.edu

Abstract: The Semiotic Pivots and Activity Spaces for Elementary Science (SPASES) Project was implemented as a proof of concept. Our goal was to demonstrate that with the right set of technological supports, young children can start their learning trajectory in science off on the right foot by engaging in rich scientific investigations into complex science topics. The SPASES curriculum was successfully implemented in two multi-age classrooms of 43 students aged 6-8 years at a progressive elementary school in Los Angeles, CA. Pre/Post-test results show that these 6-8 year old students were able to develop a conceptual understanding of force, net force, friction and two-dimensional motion after participating in the SPASES curriculum which leveraged their prior experiences and ability to engage in embodied play as a form of scientific modeling.

Introduction

Early elementary science instruction has not kept pace with the developmental literature on young students' cognitive competencies that can be used as building blocks for understanding science concepts (NRC, 2007; Metz, 1995). In fact, young children can, under the right circumstances, do more and learn more complicated ideas than we currently ask of them in early elementary science education. One argument against 'ambitious' science instruction (1) is that aspects of classical experimental design such as controlling variables and separating hypotheses from evidence have proven difficult for young children (Klahr, 2000; Schauble, 1996; Siegler & Liebert, 1975). However, alternative studies have shown that asking students to produce and evaluate models of the real world to help them generate predictions can make it possible for them to effectively participate in the process of scientific knowledge production and learn the content being studied (Lehrer & Schauble, 2006). Models—and in the case of the SPASES project, hybrid models that leverage both computer simulations and physical embodiment to describe Newtonian force and motion—are a critical part of the scientific inquiry process and can help students coordinate theory with evidence (Schwarz & White, 2005). However, while modeling is within reach of early elementary students, they still do not progress very far without carefully scaffolded experiences (Lehrer & Schauble, 2000).

In this paper we describe how first and second grade students learned about the physics of force and motion through a series of technologically enhanced modeling activities. At the heart of the project was a set of augmented reality and motion-capture technologies called SPASES (Semiotic Pivots and Activity Spaces for Elementary Science) that were used to leverage students' existing competencies in pretend play and transition them to formal and symbolic models of force and motion. Briefly (a fuller description is provided below), cameras filmed the area at the front of the classroom. The video feed was passed through object recognition software that recognized and tracked (e.g., the position and orientation) a predefined set of geometric patterns. Students held or wore these patterns as they moved about the room. A projection of the SPASES simulations software was displayed on an interactive whiteboard. The simulation software showed the video feed of the students moving around the room. The simulation software also displayed an image of the object that the students were play-acting (e.g. a ball) superimposed by the computer software over their image in the video feed. The superimposed objects would move around the projection in real-time as the students themselves also moved around the room.

Theoretical Framework and Design Principles

Young Children and the Concepts of Force and Motion

Infants develop an intuitive notion of objects, including their permanence and their properties. By preschool these intuitions have developed into a sophisticated sense of mechanical causality and understanding of the links between unseen causes and observable results (Bullock, Gelman, & Baillargeon, 1982; Yoachim & Meltzoff, 2003,). Additionally, pre-school children can distinguish between distance, speed, and time when observing objects in motion (Acredolo, Adams, & Schmid, 1984; Matsuda, 2001). Even so, some concepts of force and motion are difficult for young students to grasp and these core but challenging concepts were the target of our instructional intervention.

The first concept that SPASES focused on was force, including: the causal relationship between force and motion; the difference between force and speed; the fact that once a force ended, the speed of an effected object continued (i.e. inertia); and that impulse forces were an interaction between objects but not the objects themselves. These topics correspond to some of the key conceptual stumbling blocks to understanding force and motion (Lehrer & Schauble, 1998). Second, we focused on quantifying the relationship between force and speed (i.e., net force). Third, students investigated friction as a force. Finally, the curriculum focused on net forces in two dimensions.

Description of the SPASES Environment and Technology

There were two key components to the SPASES system: 1) an augmented reality system that used computer vision to record and display the students' physical actions and locations, and 2) software that translated this motion into a physics engine and generated a response based on the sensing data. The SPASES system used commercially available, open source forms of motion tracking and pattern recognition technologies to create an inexpensive alternative to virtual reality within the physical classroom (a 12' x 12' carpet at the front of the classroom). Motion tracked by the system could be instantly imported into a computer simulation that allowed students to model their understanding of force and motion and compare their predictions to simulated results.

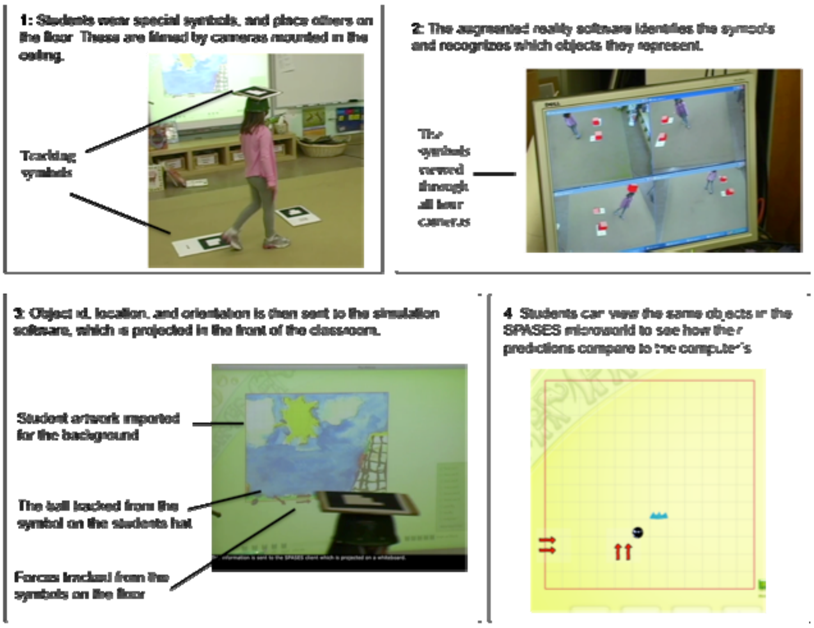


Figure 1. The Progression from Physical Objects and Motion to a Physics Microworld in SPASES.

We will describe one example activity to illustrate how the SPASES technologies supported successful modeling. In this activity the students were asked to predict how a series of forces would influence the motion of a ball. The students were split into two teams. The first team decided which forces to initially apply to a ball. The second team then chose the forces necessary to stop the ball on a given spot. The target concept was net force, addressing a common intuition that the ball would go in the direction of the last force. We expected that students holding this intuition would predict that when given a force in one direction followed by a smaller force in the opposite direction, the ball would reverse direction rather than slow down.

Susie, "playing" the role of the ball, demonstrated her prediction by walking across the rug wearing the symbol for a ball on a hat. As she walked, she responded to the forces she encountered by speeding up. The system tracked her movement in real time. While the students saw her move across the rug, they could also see a ball projected in the SPASES microworld move across the whiteboard, mimicking her movement in the physical classroom. As Susie-as-the-ball passed arrow symbols, her peers were also involved through evaluating whether they agreed with her prediction. Did she speed up and slow down in the right places? By the correct amount?

After Susie finished, the students were invited to debate her embodied prediction. They began by discussing how many forces were in each location and what their impact would be on the ball. Some students expressed common, but incomplete or inaccurate, intuitions while others shared more idiosyncratic ideas. The students then had the chance to compare their embodied predictions with the microworld that mirrored the choices they had made with the physical objects. Since the cards representing forces had already been laid on the floor as part of their activity, and because the system recognized these patterns as forces that operate in particular ways in the physics engine, all that the students had to do to see if they were right was reposition Susie-as-the-ball back to the beginning and press a button to run the simulation. Now the physics engine took over Susie's ball and displayed what would happen for that same scenario in a Newtonian world using the same space and representational system as the children's pretend play. Ultimately, everyone was surprised when their predictions did not quite match the computer simulation, and in the ensuing discussion students made explicit some of their implicit thinking. This discussion provided a key building block for a series of activities that led to the majority of the students in the group to build on what they already knew (or thought they knew) about how things move and to begin to reason in a more normative manner about how forces contribute to an objects motion. Thus the students started the activity using pretend play skills, but by the end of the activity were engaging in a discussion about modeling and concepts of net force. Through this game-like experience, SPASES makes it possible for 6-8 year-old students to interrogate their own understanding and explore these physics concepts. We now turn to the broader theoretical framework that guided our design.

Design Principle #1: Play and Participatory Modeling

For young students in particular, it is important to develop modeling abilities by starting with what they can already do. This is a fundamental premise of constructivism—that students' existing schemata are modified, added to, and reorganized, but not abandoned during the learning process (Smith, diSessa, & Roschelle, 1994). An understanding of modeling begins with symbolism, as models stand for something else and often use collections of symbols to do so. Importantly for the SPASES project, as early as pre-school, children are able to distinguish toys, pictures, and video images as representations of real objects, and can use representations successfully to reason about the world (DeLoache & Burns, 1994).

In addition to nascent symbolism, young students have another important competency at their disposal for representation—one that is not traditionally thought of as a building block for science, but which we believe can be effectively marshaled to that end—this competency is play. Play, particularly embodied, socio-dramatic play where children use their bodies and movements to enact a scene or situation, is an activity that young children are competent at and familiar with from an early age, and which is closely tied to the development of symbolic representation (Nicolopoulou, 1993; Piaget, 1952).

The defining feature of pretend play is not that it is fun (although it often is). Rather, its defining feature is the combination of an imaginary situation with a set of rules (Vygotsky, 1978). In pretend play, students are able to engage with quite complicated rule sets. For example, when "playing house," children typically control their behavior based on a set of rules about what fathers do, what mothers do, and what babies do. It is this focus on a set of rules that makes play a potential resource to help students learn science. Scientific phenomena are often described as a set of rules or laws—for example, Newton's three laws of force and motion.

The rules in pretend play are what make play a valuable part of the learning process and a type of informal inquiry (Youngquist & Pataray-Ching, 2004). In play children often attempt to govern their behavior by following a set of rules that they do not yet fully understand. Additionally, through play, the rules that govern a situation become visible and often explicit for children (Rosenberg, 1987). Understanding the rules that govern the world is one of the central aspects of scientific modeling. For this reason, researchers have argued that play is an early form of simulation (Bruner, 1986).

To incorporate play into the SPASES curriculum, we engaged students in developing and refining *participatory models* (Danish, 2009). Participatory models are embodied, dramatic skits where the students enact a key principle of the system being studied, and leverage their body motion and position as a resource for displaying their understanding. Participatory modeling builds upon the kind of productive engagement that has been seen in participatory simulations (Colella, 2000) while shifting the focus on rules to be more explicit and reflective for the participants. By identifying these play activities as participatory modeling, we are highlighting the fact that students were presenting, through embodied enactment, their model of how the ball would move.

To facilitate productive modeling throughout the curriculum, SPASES began with a first-person experience—an important building block for young students' scientific understanding—where one student pretended to be the ball and used his/her own physical motion to predict and represent the motion of the ball. It has been shown that when learning difficult science concepts, students benefit from examining the system from multiple perspectives, particularly in computationally supported environments where the technology can help students take perspectives beyond their own perceptual capabilities (c.f., White, 1993). Like traditional computer simulations, SPASES offers the outside observer's perspective as well, where one can look down from above and observe forces, friction and motion, running experiments and measuring the phenomena (see Figure 1). However, given the age of our students, SPASES began with a first-person experience and then transitioned to an abstracted third person perspective.

Design Principle #2: Progressive Symbolization

An additional intersection between play and scientific activity is the role of symbolism in play. In play, the child can choose which features of the situation are relevant and meaningful and which features can be ignored. This is exactly what children have difficulty with, when engaging in formal scientific investigations. Young students frequently insist on fidelity, especially visual fidelity, requiring that the model and representation look the same (e.g., water is blue, leaves are green, etc.). For example, a child who pretends a blue cloth is a lake that her toy boat must cross has somewhat rigidly used the similarity in color to assign a symbolic meaning to the cloth. At the same time, she has flexibly chosen to ignore other aspects of the cloth, such as its square shape and lack of wetness, and by not assigning them significance, making them semiotically invisible. Thus, in play students are able to fluently use symbolism and abstraction in ways that remain difficult for them in other contexts such as formal investigations.

In SPASES the artwork and symbols that populated the system were all invented by the students themselves. Like previous work in progressive symbolization (Enyedy, 2005), the students refined their symbols as a group, determining which aspects of the phenomenon would be captured in a symbol and iteratively refining those symbols so that they would be effective in their future modeling activities.

Design Principle #3: Cycles of Activities and Semiotic Resources

After making embodied predictions, the students seamlessly transitioned into a physics microworld to compare their embodied predictions to what would actually happen in a perfect Newtonian world. Students positioned objects within SPASES using either the shared interactive whiteboard, or the augmented reality objects. Like prior effective microworlds (c.f., White, 1993), SPASES allowed students to see and manipulate a situation in ways impossible in the real world (e.g., turning off friction). Asking students to place objects on the whiteboard or in the physical classroom had the added benefit of creating public and open tools for discussion (Danish & Enyedy, 2007; Hutchins, 1993). This openness was an important feature of SPASES, allowing students to interrogate their peer's choices or propose alternative predictions for what they thought would happen.

However, students also engaged in non-computer-mediated experiences and investigations in the real world, as well as play-acting without technology, or technology without pretend play and tracking. This range of activities was intended to connect student understandings at multiple levels of abstraction—from actual balls they could touch to symbols about motion devoid of any reference to the objects doing the moving.

Methods

Participants

The SPASES curriculum was successfully implemented in two multi-age classrooms with students aged 6-8 years (x=7.1 years) at the UCLA Lab School (n=43). The forty-three students were roughly even in terms of first and second grade students (Twenty two 1st graders & twenty one 2nd graders) and in terms of gender (21 boys & 22 girls). The ethnicity of the children roughly mirrors the ethnicity of the state of California (although Latinos are under-represented in our sample); 53% Caucasian, 22% African American, 14% Latino and 11% Asian.

The curriculum lasted 15 weeks (2/18/09 through 6/8/09) and consisted of 26 one to two hour sessions. The average length of a lesson was 90 minutes. Four major topics were covered; force and speed (5 lessons), net force in one dimension (11 lessons), friction (4 lessons), and two-dimensional motion (7 lessons).

Procedures

Students were individually interviewed before and after the unit with a protocol based in part on a modified version of the Force Concept Inventory (Hestenes, Wells, & Swackhamer, 1992). To document learning processes and how the curriculum was enacted by the teachers, we videotaped two case study groups (students were organized into small groups of 8-9 students) and all whole-class activities.

The pre- and post-test interviews were transcribed and coded for degree of conceptual understanding. Reliability for each item was determined by calculating the Intra-class Correlation Coefficients (ICC) for each item. Five of the 34 items were dropped because of low inter-rater reliability. An additional ten items were dropped due to a high proportion of missing answers. These missing answers were due in part to student attendance, but also due to variability in the way that various members of the research team administered the interview, and the difficulty in parsing the continuous transcript into discrete answers. As a result, the final pre-test and post-test scales were comprised of nineteen items. Reliability analyses were conducted on the pre-test and post-test items to ensure that the data had a unidimensional structure. The Cronbach's alpha for the *pre-test scale* was .34. The Cronbach's alpha for the post-test scale was .57. Two explanations may account for the low alpha values. First, the sample size was small; with a larger sample reliability estimates are expected to be higher. Also, our participants were young children who had very little knowledge of the concepts prior to the study. For example one student thought acceleration was a type of vegetable. Therefore, the low consistency among items on the pre-test may be an artifact of their age and understanding, rather than poor inter-item correlation. This assertion is supported by the dramatic increase in reliability after the intervention.

Results

Descriptive statistics were obtained on performance on the pre-test and post-test items. For the 43 students, the average pre-test score was 5.42 (*SD* = 1.55) out of a possible of nineteen points. The average post-test score was 8.60 (SD = 2.07). A paired-samples t-test was conducted to compare pre-test scores and post-test scores. Post-test scores were significantly higher than the pre-test scores, $t(42)= 10.43$, $p < .005$. Correlational analyses examined the relation between grade level, age at the start of the study, gender, pre-test and post-test scores. Results indicate there is no correlation between any of the demographic variables and the assessment scores (see Table 1).

Scales were formed for our four content objectives: *Force and speed, friction, net forces,* and *two-dimensional motion.* However, due to the small number of items per scale, the reliability of each scale was extremely low. Therefore, in order to examine differences in content understanding on these four specific topics, we have analyzed four exemplary questions. A Wilcoxon signed rank test was computed to examine changes in scores on each of these items. The Wilcoxon signed rank sum test is a non-parametric version of a paired sample t-test, which we chose to use because it requires fewer assumptions about the distribution of the data.

For the topic of force and speed, we analyzed the question that asked students "What is a force?" The highest value was given to answers that reflect the understanding that either force makes something go proportionately faster or slower, or that forces change the speed of an object. Partial credit was given to answers that describe forces as a verb (e.g., it makes something move) or as a noun (i.e, provides an example of a force). The sign test indicated that 28 (65%) of the students received higher scores on the post-test than on the pre-test, $Z = 4.83$, $p < .005$.

For the topic of friction, we analyzed responses to a scenario that asked students to explain why a moving soccer ball slows down when rolling on a grassy surface. The highest value was given to students who described the resulting action and the mechanism of the friction (e.g., "Because those things sticking out of it, it will hold them back, it will try to push the ball back and stop."). Partial credit was given to answers that either described the surface quality of the grass (e.g., "So that's why it slows on the grass, because it's a little bumpy.") or connected the change in speed to friction or the grass (e.g., "Because it's really high friction right here, that's where it stops."). The sign test indicated that 16 (37%) of the students received higher scores on this question during the post-test than on the pre-test, although the results were marginally significant, $p = .052$.

Table 1: Pearson correlations between background variables and test scores.

	Pre-test	Post-test	Age at start	Grade	Gender
Pre-test	1.00	.416**	0.26	0.21	-0.14
		0.01	0.09	0.16	0.38
Post-test	.416**	1.00	0.11	0.19	0.05
	0.01		0.48	0.23	0.75
Age at start	0.26	0.11	1.00	.835**	0.15
	0.09	0.48		0.00	0.35
Grade	0.21	0.19	.835**	1.00	0.23
	0.16	0.23	0.00		0.14
Gender	-0.14	0.05	0.15	0.23	1.00
	0.38	0.75	0.35	0.14	

For the topic of net forces, we analyzed responses to the questions "What size force would you give to stop a ball that got the large size force? Why would you do that?" The highest value was given to responses that provided the correct amount of force (i.e., the same amount of force) and explained that an equal number of forces must be applied in order to stop an object (e.g., "Same force hitting each other would probably just stop." Partial credit was given to students who simply provided the solution but no explanation. The sign test indicated that 14 (33%) of the students received higher scores on this question during the post-test than on the pre-test, although the results were marginally significant, $p = .12$.

For the topic of two-dimensional motion, we analyzed the response to the modified FCI item that asked students to predict the path of a puck that received another hit (see Figure 2). The sign test indicated that 29 (67%) of the students got higher scores on the post-test than on the pre-test, $Z = 4.67$, $p < .005$.

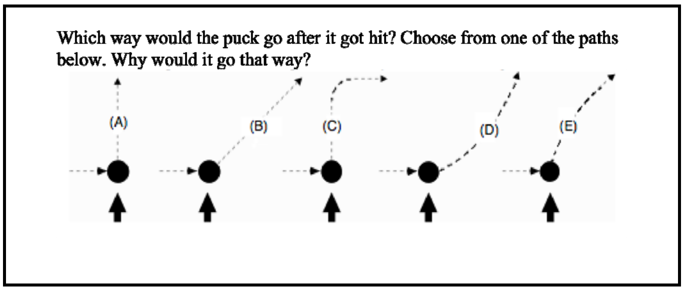


Figure 2. Modified FCI Question about 2-D Motion.

In sum, students demonstrated significant improvement on all of the key measures, including, the fact that 91% of the students showed a pre- to post-test gain ($Z = 5.71$, $p < .005$).

Conclusion

SPASES is an important proof of concept project. We aimed to demonstrate that young children can begin their learning trajectory in science off on the right foot—both in terms of the complexity of science content and the type of ambitious science instruction that will lead to generative inquiry skills and a robust scientific epistemology. Pre/Post-test results show that these 6-8 year old students were able to develop a conceptual understanding of speed, force, friction and two-dimensional motion.

What we have shown here is that the students are able, with the SPASES technology and activities to learn force and motion concepts at an earlier age than thought possible.

We were pleased to see that neither gender nor age were correlated to post-test performance. We were initially concerned that the SPASES environment might appeal more to and therefore provide a greater benefit for boys. The environment overlaps with many of the stereotypical interests and styles of boys'—it involves a mechanical topic, involves physical activity, and heavily depends on computer simulations and gaming. Nevertheless, from our videotapes we saw that girls were just as deeply engaged during the activities as boys and contributed substantially, if not to a greater extent, during the whole-class and small group discussions.

We were particularly surprised by two of our findings. While our overall results were encouraging, the sub-topic results showed some unevenness in student learning. We had relatively small gains in students' ability to quantify the relationship between speed and distance and their understanding of friction. In contrast, we had relatively large gains in students understanding of two-dimensional motion, a topic that has proven difficult for much older students.

With regards to friction, much of the students' difficulty can be traced back to two factors. First, students came in with more experience with friction both in and out of the classroom, and thus scored higher on the pre-test on these items. Second, students' intuitions conflicted with our example of ice as a low friction environment. As stated above in our third design principle, we were committed to having some sort of physical and familiar environment for students to be able to explore. Given this commitment, we had relatively few inexpensive options of familiar non-friction/low friction environments—air hockey tables and ice. Neither was ideal in that both introduced new mechanisms (an upward force and lubrication respectively). We choose ice on the assessment (and oiled surfaces as an alternate to ice in the activities) because the net balance between gravity and the upward air pressure in the hockey table seemed to necessitate a discussion of gravity—a topic that was not covered by our curriculum. Perhaps because our dramatic play activities were kinesthetic in nature, we found that a large number of students were bringing in their memories of falling on ice, and the sensation of their legs speeding up as they fell. As a result, students inferred that in no/low friction environments, objects sped up rather than maintained their inertia. This interpretation of their past experience interacted with our activities in unanticipated ways, contributing to our weaker results on this topic.

The results for two-dimensional motion, however, surprised us for the opposite reason. Given how entrenched the intuition is that an object will travel in the direction of its last hit, and the difficulty that older students have shown on this FCI assessment item, we had modest expectations for this topic in our curriculum. While the majority of our students at the time of the post-test were limited to a qualitative sense of the direction and speed of the new vector, we were encouraged that our results were similar to the results obtained by White's (1993) seventh grade students after the Thinker Tools software and curriculum. Based on our preliminary analysis of the video records, we attribute the students' success in this area to the additional semiotic resources the students had in the augmented reality environment. Further, the ways in which embodied action was annotated and formalized helped to create what others have called semiotic fusion (Nemirovsky, 2003), liminal spaces (Ochs, Gonzales, & Jacoby, 1996) and conceptual blends (Fauconnier & Turner, 1998). In our case, embodied actions laminated with symbol systems invented by the students were used as a key resource to ground abstract aspects of the students' models of force and motion. This line of reasoning warrants future study, as it is at the heart of the question of why the SPASES environment worked and would help determine what might generalize from this study to other studies and other computer-mediated environments.

Endnotes

(1) We have adapted the term 'ambitious math instruction' (Lampert, Beasley, Ghousseini, Kazemi, & Franke, 2010) which was used to refer to instruction that simultaneously targets conceptual understanding, procedural fluency and productive dispositions towards the domain.

References

Acredolo, C., Adams, A., & Schmid, J. (1984). On the understanding of the relationships between speed, duration, and distance. *Child Development, 55*, 2151-2159.

Bruner, J. (1986). *Actual Minds, Possible Worlds*. Cambridge, Mass.: Harvard University Press.

Bullock, M., Gelman, R., & Baillargeon, R. (1982). The development of causal reasoning. In W. J. Friedman (Ed.), *The developmental psychology of time*. New York: Academic Press.

Colella, V. (2000). Participatory simulations: Building collaborative understanding through immersive dynamic modeling. *Journal of the Learning Sciences, 9*(4), 471-500.

Council, N. R. (2007). *Taking Science to School: Learning and Teaching Science in Grades K-8*. Washington, DC: National Academies Press.

Danish, J. A. (2009). *BeeSign: A computationally-mediated intervention to examine K-1 students' representational activities in the context of teaching complex systems concepts.* Unpublished PhD thesis, University of California.

Danish, J. and Enyedy, N. (2007). Remember, We Have to do all the Parts of the Rose: Negotiated Representational Mediators in a K-1 Science Classroom. *Science Education* 91:1-35.

DeLoache, J. S., & Burns, N. M. (1994). Early understanding of the representational function of pictures. *Cognition, 2*(52), 83-110.

Enyedy, N. (2005). Inventing Mapping: Creating cultural forms to solve collective problems. *Cognition and Instruction* 23(4), 427 - 466.

Fauconnier, G., & Turner, M. (1998). Conceptual integration networks. *Cognitive Science, 22*(2), 133-187.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory.

Hutchins, E. (1993). Learning to Navigate. In S. Chaiklin & J. Lave (Eds.), *Understanding Practice: Perspectives on Activity and Context* (pp. 35-63). Cambridge: Cambridge University Press.

Klahr, D. (2000). *Exploring science: The cognition and development of discovery processes*. Cambridge, MA: MIT Press.

Lampert, M., Beasley, H., Ghousseini, H., Kazemi, E., & Franke, M. (2010). Using designed instructional activities to enable novices to manage ambitious mathematics teaching. *Instructional Explanations in the Disciplines*, 129-141.

Lehrer, R., & Schauble, L. (1998). Reasoning about structure and function: Children's conceptions of gears. *Journal of Research in Science Teaching, 31*(1), 3-25.

Lehrer, R., & Schauble, L. (2000). Inventing data structures for representational purposes: Elementary grade students' classification models. *Mathematical Thinking and Learning, 2*, 49-72.

Lehrer, R., & Schauble, L. (2006). Scientific thinking and science literacy. In R. W. Damon, K. Lerner, A. Renninger & I. E. Sigel (Eds.), *Handbook of child psychology, 6th edition, (vol. 4)*. Hoboken, NJ: Wiley.

Matsuda, F. (2001). Development of concepts of interrelationships among duration, distance, and speed. *International Journal of Behavioral Development, 25*, 466-480.

Metz, K. E. (1995). Reassessment of Developmental Constraints on Children's Science Instruction. *Review of Educational Research, 65*(2), 93-127.

Nemirovsky, R. (2003). *Three Conjectures concerning the Relationship between Body Activity and Understanding Mathematics.* Paper presented at the PME 27, Honolulu, Hawai-I.

Nicolopoulou, A. (1993). Play, cognitive development, and the social world: Piaget, Vygotsky, and beyond. *Human Development 36*(1), 1-23.

Ochs, E., Gonzales, P., & Jacoby, S. (1996). "When I come down, I'm in a domain state": Grammar and Graphic Representation in the Interpretive Activity of Physicists. In E. Ochs, E. A. Schegloff & S. Thompson (Eds.), *Interaction and Grammar* (pp. 328-369). Cambridge: Cambridge University Press.

Piaget, J. (1952). *Play, dreams and imitation in childhood*. New York, NY, US: W. W. Norton & Co, Inc.

Rosenberg, H. S. (1987). *Creative Drama and Imagination*. New York, NY: Holt, Rinehart & Winston.

Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology, 32*(1), 102-119.

Schwarz, C., & White, B. (2005). Meta-modeling knowledge: Developing students' understanding of scientific modeling. *Cognition and Instruction, 23*(2), 165-205.

Siegler, R. S., & Liebert, R. M. (1975). Acquisition of formal experiment. *Developmental Psychology, 11*, 401-412.
Smith, J. P., diSessa, A. A., & Roschelle, J. (1994). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *Journal of the Learning Sciences, 3*(2), 115–163-115–163.
Vygotsky, L. S. (1978). *Mind in society : the development of higher psychological processes*. Cambridge: Harvard University Press.
White, B. (1993). Thinker tools: Causal models, conceptual change, and science education. *Cognition and Instruction, 10*(1), 1-100.
Yoachim, C. M., & Meltzoff, A. N. (2003, October). *Cause and effect in the mind of the preschool child*. Paper presented at the biennial meeting of the Cognitive Development Society.
Youngquist, J., & Pataray-Ching, J. (2004). Revisiting "Play": Analyzing and Articulating Acts of Inquiry *Early Childhood Education Journal 31*, 171-178.

Group Sense Making of Multiple Sources in a Hypertext Environment

Sarah A. Sullivan, Katherine D. Knight, Sadhana Puntambekar, Department of Educational Psychology, University of Wisconsin, Madison, WI
Email: sasullivan2@wisc.edu, kknight@wisc.edu, puntambekar@education.wisc.edu

Abstract: Collaboration is important for learning science through inquiry. Hypertext and hypermedia environments are increasingly being used to help support inquiry and enable students to build background conceptual knowledge. Previous research has found that students have difficulty reading multiple texts in groups, so this study explored collaborative discussion of multiple sources in a hypertext system after students read individually. Results indicated that the group that engaged most frequently in monitoring the understanding of group members, evaluating ideas and generating questions to share with the class developed the most sophisticated scientific conceptions. Simply navigating the hypertext system effectively and engaging in elaborated science reasoning was not sufficient for students to develop sophisticated conceptual understandings, reiterating the importance of helping young learners to actively reflect on their understanding and question development during collaboration.

Introduction

Inquiry-based learning is an increasingly important part of many science classrooms. Learning science through inquiry requires collaboration to make sense of the multiple forms of information that students are asked to reason about and integrate in order to learn about scientific concepts (NRC, 1996; Olson & Loucks-Horsley, 2000; Varelas & Pappas, 2006). One resource that students are often asked to make sense of is scientific texts, especially digital text in the form of hypertext and hypermedia systems. Hypertext and hypermedia environments can be an effective way to present multiple sources of information as well as support learning by making relationships among information sources visible through the structure of the system (e.g., Rouet, Potelle, & Goumi, 2005; Gerjets, Scheiter, Opfermann, Hesse, & Eysink, 2009). Including digital texts as part of the inquiry process can help students build the background conceptual knowledge that they need to supplement their in-the-moment experience with an understanding of the concepts underlying their activities. Students in inquiry classrooms are frequently required to reason about multiple texts. But even though reading is largely an individual activity, students are often asked to read and make sense of these multiple texts in collaborative groups. Previous studies that have investigated groups collaboratively reading multiple texts have identified problems with this arrangement, such as a lack of active engagement by the entire group with the texts (Schwarz, 2003) and a lack of engagement in evaluation of the text resources and text content while reading collaboratively (Wiley & Bailey, 2006). However, since collaboration is an essential part of learning science through inquiry, more research is needed on how students can productively work together to actively make sense of multiple textual sources to support their inquiry activities.

Facilitating active engagement when thinking about scientific concepts is important because relationships are very often complex, and learners may have difficulty understanding conceptual relationships. For example, common to the findings of researchers investigating momentum and impulse is that students lack a coherent understanding of momentum, confuse the scientific terms of momentum and impulse with their meanings in everyday language, and are very poor at making causal connections between momentum and impulse (e.g., Bryce & MacMillan, 2009; Lawson & McDermott, 1987). A recent review by Nussbaum (2008) concluded that collaborative group discourse can enhance conceptual understanding of content when diverse student views of conceptual principles are considered and evaluated and group members model elaborative and metacognitive strategies. Additionally, research has found that groups in which students facilitated members' comprehension of content by asking for explanations and challenging each other's conceptual understanding showed greater learning gains than groups that only engaged in explanation construction (e.g., Asterhan, & Schwarz, 2009).

One potential solution to help facilitate engagement with texts and still allow time for the collaborative discussion important to inquiry is to have students read the texts individually and then come together as a group to make sense of what each student has learned. In a previous study (Smith, Sullivan, & Puntambekar, 2009), we found that students who navigated in a hypertext environment individually and then came together to discuss were less off-topic and shared more information about science concepts than students who discussed content while they were navigating as a group. Therefore, we conducted this study to examine how students collaboratively made sense of information presented in multiple documents in the hypertext system that they had read individually to prepare for their hands-on inquiry exploration. Our goal was to examine the kinds of discourse that seemed to help students to successfully develop a more sophisticated understanding of physics concepts and what kinds of support could be provided either by the system or the teacher to improve their discourse and learning.

Methods

The data were collected in an 8th grade science classroom in which students used the hypertext system CoMPASS (Puntambekar, 2006) as part of a science unit on Forces and Motion (see Figure 1). Students used CoMPASS to learn about concepts related to forces and motion in order to prepare for five design explorations throughout the unit. The goal of the unit is to design roller coasters and other rides for an amusement park. Prior to the Forces and Motion unit, students completed a unit on Work and Energy, in which they also used the hypertext system CoMPASS. Thus this was at least students' eighth experience using the CoMPASS system, and they were therefore familiar with its navigation and layout. For this study, students navigated CoMPASS individually for approximately 30 minutes to research concepts for the Stop the Car exploration, which required them to figure out how to stop their roller coaster car safely by applying concepts such as *momentum* and *impulse* to their design. For this exploration, students should have focused on concepts within the topic of linear motion. After navigating individually, students were given approximately 10 minutes to meet in their groups and discuss what they had learned about the physics concepts for their exploration. The class in which data were collected consisted of five groups with three students in each group. The teacher determined group composition based on students she felt could collaborate productively. Groups were named after elements: Helium, Oxygen, Argon, Sodium, and Zinc, and consisted of both boys and girls. In addition to discussing how concepts would inform the design for their hands-on exploration, another group goal was to develop questions about concepts they didn't understand in order to prepare for a whole class discussion.

The CoMPASS hypertext system was developed as part of a design-based physics curriculum in which students engage in design exploration challenges and investigations. CoMPASS provides navigable concept maps that were designed by physics experts to mirror the conceptual structure in physics, affording students a visual representation to help them gain a rich understanding of concepts and their numerous relationships. The CoMPASS system provides two representations: a navigable concept map and text that describes the concepts. The concept that the student selects to read about becomes the center (focal) point of the map and the other concepts move accordingly based on the strength of their relationship to the center concept.

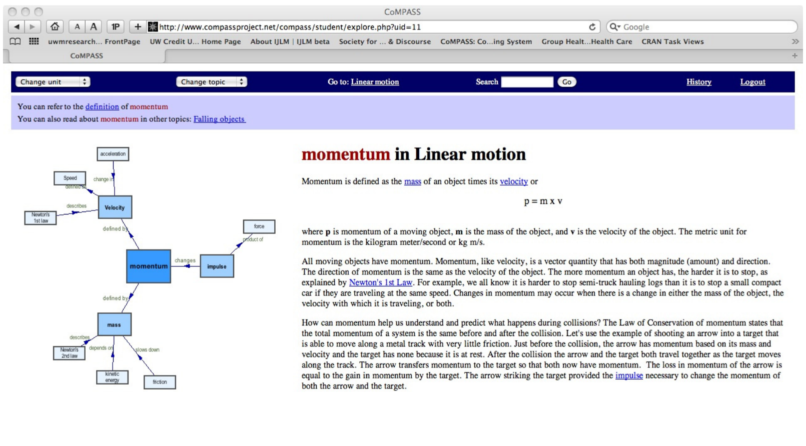


Figure 1. Screenshot of the CoMPASS Page for the Concept of Momentum in the Topic of Linear Motion.

Data Sources and Analysis

For this study, we used multiple sources of data from five groups of 8th grade students. We looked at (i) students' individual initial conceptions of the concept of impulse using a pretest (described below) that was given before they began the unit, (ii) patterns of navigation exhibited by each of the group members while using CoMPASS, (iii) group audio of students' dialogue as they attempted to make sense of the information from the multiple texts in CoMPASS, and (iv) students' conceptions of impulse at the end of the unit after the Stop the Car exploration using the same question as the pretest.

Pre and Posttest Conceptions of Impulse

As part of the pre/posttest, students were given an open-ended essay question about impulse. The purpose of the question was to elicit students' initial conceptions about impulse and to see whether they developed a more sophisticated understanding of impulse throughout the unit. Students were required to write a short paragraph addressing the following scenario:

> Imagine throwing two identical eggs with the same amount of force. The first egg is thrown against a cement wall. The second egg is thrown into a deep pile of hay. The egg thrown into the wall breaks, but the egg thrown into the pile of hay does not. Please use science concepts *impulse* and *momentum* to explain the different results.

Pre and post-test responses were scored using a 0 to 5 point scale. Categories were drawn from research on students' understanding of momentum and impulse (e.g., Bryce & MacMillan, 2009; Lawson & McDermott, 1987) that examined common initial conceptions, especially the difficulty with which students make causal connections between momentum and impulse. The increasing scale represents student responses that suggest differential understanding beginning with everyday language, inclusion of appropriate science concepts, recognition of a causal relationship, and finally the identification of a mechanism within the relationship. For example, a level-1 response would be, "The egg landing in the hay has soft impact compared to the one thrown into the cement wall." Whereas a level-5 response would be exemplified by:

> Both eggs have the same mass and velocity, so they have the same momentum. But, since the cement wall has very little give, the egg hits it with a greater force because the time isn't spread out. But the egg which hits the hay doesn't break because although both eggs hit with the same impulse, the hay extends the time it takes for the egg to stop which decreases the force that it hits with.

We combined the scores for all three of the group members in order to obtain a score that represented the level of understanding of impulse for the group. We did this because we wanted to focus on development of understanding by the group as a whole. A group's score on the question could range from 0 to 15. Using these scores we broke the groups down by their level of understanding of impulse: Low 0 to 5, Medium 6 to 10, and High 11 to 15 points.

CoMPASS Navigation

Computer log files were used to look at students' navigation behavior, while they used CoMPASS to conduct research for their hands-on exploration, in order to examine whether navigation patterns were related to group dialogue and understanding of impulse. The log files recorded in chronological order the topics and concepts that students visited and the time spent on each. In order to understand student navigation in relation to the goal of the hands-on exploration, log files for the Stop the Car exploration were analyzed using Pathfinder (Schvaneveldt, 1990). Pathfinder is a graph theoretic technique used to create network representations consisting of nodes and links that characterize a group's navigation patterns in CoMPASS, allowing us to look at similarities and differences in navigation paths and whether students missed key information. The graphics of the navigation patterns (see Figure 2) show the number of times students "navigated to" a concept and the number of times students "navigated from" a concept and from which concepts students navigated to others. For example [3:2] means that a student "went to" a concept three times and "went from" that concept to another concept two times. Navigation patterns make visible the choices that a reader makes in following links, and help us to understand the process with which a reader constructs the meaning of the text (Bolter, 1998). We combined the navigation patterns of all three of the members in each group in order to obtain a complete representation of all concepts visited by students in the groups.

Analysis of Student Discourse

We transcribed the audio of the student discourse during the time that students came together in their small groups to discuss what they had learned on CoMPASS to help them with their Stop the Car exploration. We were interested in understanding the types of student discourse that helped students to actively think about their understandings of the science concepts in order to prepare them for their whole class discussion and hands-on exploration. Total audio data consisted of approximately one hour of audio and 33 pages of transcripts. A coding rubric was developed based on prior work examining students' collaborative talk when using CoMPASS in the classroom (e.g., Bopardikar, Sullivan, & Puntambekar, 2009). A process of axial coding (Strauss & Corbin, 1998) was then employed whereby the existing coding rubric was applied to the transcripts and modified based on the types of discourse that emerged from the initial analysis of the transcripts, resulting in eight codes (see Table 1).

We coded the transcripts at the utterance level, which we defined as an individual turn at talk by a student. All groups had the same amount of time to discuss the content (10 minutes), but varied in the degree to which they were on task and focused on physics ideas. As a result, transcripts were reviewed to identify the areas in which students were discussing physics concepts, and only these sections of the transcripts were coded. Utterances could be multiply coded, and utterances that could not be assigned a code were eliminated. The first two authors applied the coding rubric to 20% of the transcripts and established an interrater reliability of 90%. Disagreements were resolved through discussion. The first author coded the remaining transcripts. We then counted the number of times that the various types of talk were coded for each group.

Table 1: Codes used for student discourse.

Code:	Definition:	Examples:
SE: Surface Science Explanations	Stating a definition, formula, or science idea without elaborated explanation. This includes just reading from notes.	“It says...well it said something...about ((reading)) law of conservation states that total momentum is the same before and after a collision.”
ER: Elaborated Science Reasoning	Going beyond stating a definition or formula to elaborately explain ideas or concepts with discourse related to interpretation, analysis, evaluation, reasoning, and problem-solving skills employed to think through science ideas and concepts. For example, using concrete explanations or examples to better understand the abstract concepts in CoMPASS.	“That there's a...there's a -- there isn't velocity, so there shouldn't be any momentum.” “Yeah 'cause the oil...like, [you] grease it up before and... it reduces the friction.”
SM: Science Misconceptions	Making an incorrect statement about science concepts.	“If you have impulse, you have momentum. If you don't have impulse, you don't have momentum.”
QR: Question Raising	Questions raised by group members about science concepts for group discussion.	“Would friction cause brownies to get stuck to a pan?”
RC: Reference to Information from CoMPASS	Reference to specific content about concepts coming from CoMPASS and discussion of what information was read or found by group members.	“It says online -- er, on the thing that it was a force.” “Or wha -- like...yeah, did anybody look up anything on Newton's laws? I didn't.”
RE: Reference to Current Exploration	Reference to the exploration for which they are using CoMPASS to investigate concepts.	“That would help us with the stop the car exploration because it's like...”
CD: Ideas for Class Discussion	Reference to ideas or questions for class discussion.	“Um, maybe we can do that in the class discussion when …”
GM: Group meaning making processes	Recognizing a personal lack of understanding about an idea or concept and sharing this with the group. Checking the understanding of other group members. Challenging a science idea articulated by another student.	“I don't understand how friction can be a force.” “Okay. Do you understand it now [S3]?” “But it couldn't push it by itself.”

Results

We will first address the changes in students' conceptions of impulse from the pre to the posttest by looking at the group score on the impulse question. We will then attempt to relate the change, or lack of, in students' understanding of impulse to their discourse when discussing in their small groups what they had learned from the multiple texts in CoMPASS while navigating individually.

Prior Knowledge and Final Conceptions of Impulse

Groups' scores on the impulse question could range from 0 to 15. We will discuss groups by their level of understanding of impulse: Low 0 to 5, Medium 6 to 10, and High 11 to 15. In terms of students' prior conceptions of impulse, the Helium and Oxygen groups both started out at a low level conceptual understanding. The Sodium, Zinc, and Argon groups all started out with a middle level conception of impulse (see Table 2).

The only group in which all students achieved a high level of understanding of impulse on the posttest was the Helium group. In fact, all students in this group achieved a level 5 score on their responses to the impulse question, which was the highest score possible. The Oxygen group remained at a low level of understanding, and Sodium, Zinc, and Argon all remained at a middle level conception of impulse, with only a slight increase or decrease in score for these four groups (see Table 2). Only one student in each of these four groups made learning gains from pre to posttest, and the other two group members either made no gain or decreased in score.

Table 2: Group totals for students' pre and posttest conceptions of impulse.

Group	Pre Total	Post Total
Helium	4	15
Oxygen	2	5
Argon	10	8
Sodium	8	9
Zinc	8	9

Group Navigation Patterns

The differences in navigation patterns among the groups can be illustrated by looking at the graphics representing navigation patterns of students in the groups (see Figure 2).

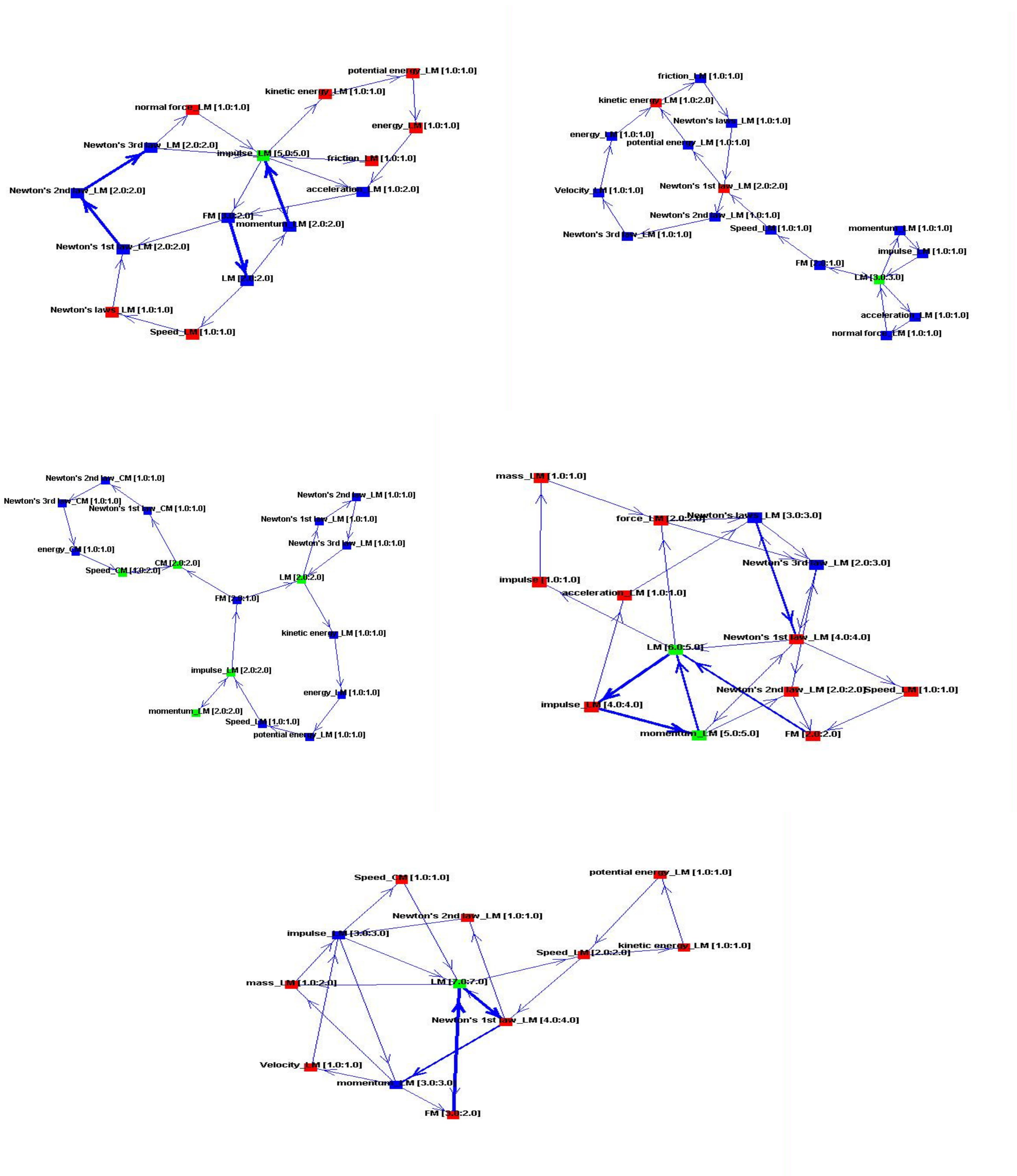


Figure 2. Combined Navigation Patterns for Group Members in Helium, Oxygen, Sodium, Zinc and Argon Groups.

These graphics represent all of the different texts about various concepts that groups visited and which, therefore, had the potential to be brought up in the group discussion. For members of the Helium group, navigation within the topic of linear motion primarily focused on impulse and transitions between impulse and momentum. Darker lines in the graphics indicate multiple transitions among concepts. Impulse and momentum are two highly related concepts that students needed to understand and be able to apply to the design for their Stop the Car exploration (see Figure 2). In contrast, members of the Oxygen group did not focus on any particular concepts and visited most of the concepts in linear motion only one time, rather than focusing on concepts they needed to understand to complete the exploration (see Figure 2). Since these graphics represent combined navigation patterns of all students in a group, this means that only one member of Oxygen visited the concepts of momentum and impulse, which were essential concepts for their exploration. The navigation patterns for members of the other three groups, Sodium, Zinc and Argon, showed varying degrees of focus on the concepts that were important for the challenge, including momentum and impulse (see Figure 2). Nevertheless, none of these three groups' navigation behaviors showed only one visit to the concepts of momentum and impulse, as did that of the Oxygen group.

Group Discourse

As explained above, there were both boys and girls in each group of three students, and the goal of meeting in small groups was to discuss how concepts informed students' design for their hands-on exploration and to develop questions about concepts that they didn't understand for a whole class discussion. There were differences in the types of talk that occurred in each of the groups (see Figure 3).

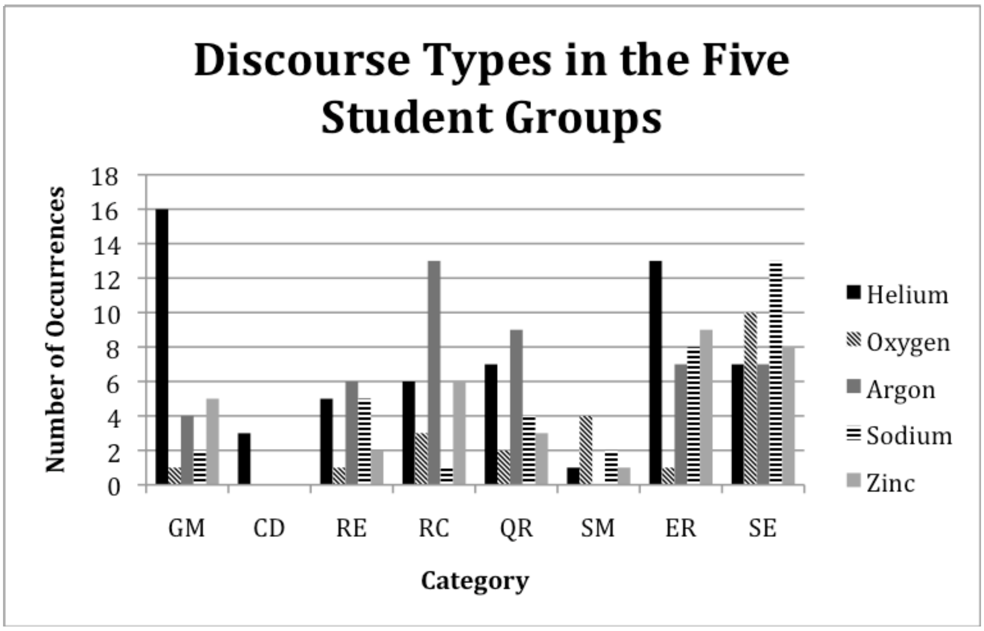


Figure 3. Number of Occurrences of Different Discourse Types in Each Group.

Figure 3 represents the number of utterances that were coded with each of the eight identified discourse types for each of the five groups. Viewing the comparison of the groups' discourse in this way reveals some interesting and significant patterns. In summary, the Helium group members differed from the other groups primarily in their much more frequent use of group meaning making discourse (GM) and in that they were the only group to develop questions for the whole class discussion (CD). To reiterate, the group meaning making code (GM) focused on recognizing one's own lack of understanding of a concept and sharing this with the group, checking the understanding of other group members and challenging science ideas articulated by other students. The Argon, Sodium and Zinc groups all had various positive exchanges, such as elaborated science reasoning (ER), talking about the application of concepts to the current exploration (RE), and question raising (QR). Nevertheless, the members of these groups lacked discourse in which they identified any lack of understanding in themselves or other group members, challenged ideas, or identified specific questions to raise in their whole class discussion. The Oxygen group only shared the information that they read on CoMPASS at a surface level (SE) without elaborating on the science and had many misconceptions (SM) without being aware of any lack of understanding.

Discussion

In this study, we examined how students collaboratively made sense of information presented in multiple documents in a hypertext system to learn about science concepts in preparation for their hands-on inquiry. Our goal was to examine the kinds of discourse that appeared to help students develop a more sophisticated understanding of physics concepts and determine what kinds of support could be provided to improve their discourse and learning.

The many instances of group meaning making (GM) by the Helium group, as evidenced by challenges to one another's ideas as well as an awareness of members' incomplete understanding of concepts in order to develop questions for the class discussion (CD), seemed to support the development of their conceptions of impulse. This outcome aligns with the findings of Asterhan & Schwarz (2009), in that students in the group that challenged and questioned their conceptual understandings showed greater learning gains than groups that only engaged in explanation construction. Further, this group also navigated among goal-related concepts in the hypertext system (see Figure 2). As described above, the Helium group was also the only group to move from a low to a high level conception of impulse. In fact, all three group members received the maximum score on the impulse question. Conversely, the Argon, Sodium and Zinc groups all remained at a middle level understanding of impulse, despite the fact that these three groups navigated to goal-related concepts (see Figure 2), which may have been facilitated by the representation of the relationships among concepts made visible by the concept map (Rouet, Potelle, & Goumi, 2005; Gerjets, Scheiter, Opfermann, Hesse, & Eysink, 2009). Additionally, these students engaged in elaborated science reasoning (ER) throughout their group conversations. However, our results indicate that although students from these three groups navigated to goal-relevant concepts and engaged in elaborated science reasoning, their discourse showed a lack of evaluation of understanding and question development by the group and did not appear to help students form a more sophisticated understanding of impulse. In contrast to the other groups, the Oxygen group started out with and remained at a low level conception of impulse. The Oxygen group did not focus on goal-related concepts (see Figure 2) during their navigation and group members focused primarily on reporting surface information (SE) from individual texts in the system. This group did not elaborate on conceptual relationships or evaluate alternate conceptions and misconceptions about the concepts. Therefore, according to the review of effective collaboration by Nussbaum (2008), it is not surprising that this group did not make substantial learning gains given that they did not evaluate student ideas and used few elaborative and metacognitive strategies. Although we only had five groups, and thus cannot make causal claims, the results indicate that conceptual understanding may be facilitated by helping students to not only just talk science but also by supporting them in helping each other to evaluate understandings and collaboratively develop questions with an explicit connection to the goal.

Implications for Practice

The results of this study suggest that the ability to evaluate what they knew and use metacognitive processes to assess conceptual understandings before going into the hands-on exploration seemed to help students develop a more advanced conception of impulse. Although the structure of the hypertext environment may have helped students to be aware of conceptual relationships, most groups could have benefited from support of evaluative and metacognitive monitoring strategies to help them think about which concepts they did and did not understand. This finding is consistent with that of Wiley and Bailey (2006) in that most groups exhibited a lack of engagement in evaluation of the text content. Mercer and Littleton (2007) raise a salient point that students often "work *in* groups but not *as* groups" (p. 26), emphasizing that students are focused on individual activities rather than working together on the task at hand. They further add that students are seldom taught how to engage in productive discourse and are often asked to "discuss this in your group" (p. 67). According to them, students need to be supported for productive collaboration. One way to support this productive discourse may be through prompts in the hypertext or hypermedia system, another through norms for discourse, such as those suggested by Mercer and Littleton, and third, through teacher support of collaboration by prompting evaluation, monitoring, elaboration and question development in group discourse. When integrating multiple digital or traditional text-based resources into their inquiry units, teachers should focus not only on helping students to explain what they have learned but also place just as much emphasis on supporting students to help them develop questions about what they don't understand.

Although having students read individually before meeting with their groups may facilitate increased discussion of scientific concepts (e.g., Smith, Sullivan, & Puntambekar, 2009), students attempting to make sense of multiple documents while collaboratively conducting scientific inquiry need help to monitor and evaluate their understanding of the text content. They also need support to actively challenge others' ideas and conceptions, and use metacognitive strategies to successfully identify questions that can then be shared with the class as part of the inquiry process. While our coding rubric allowed us to identify that these kinds of group meaning making and question development processes appear to be important to developing conceptual understanding, it only allowed us to identify these processes at the level of the individual utterance. However, future research will involve the identification of these kinds of discourse processes in more groups and look at the relationships among these processes within the group collaboration as a whole as well as identify relationships among the contributions of individual students. Since collaboration is an essential part of learning through inquiry, supporting groups to engage in discourse that stresses evaluation, monitoring, and question development may help them make connections among activities by encouraging them to actively think about what they are learning from each part of the inquiry process.

References

Asterhan, C. S. C., & Schwarz, B. B. (2009). Argumentation and explanation in conceptual change: indications from protocol analysis of peer-to-peer dialog. *Cognitive Science, 33*, 374-400.

Bolter, J. D. (1998). Hypertext and the question of visual literacy. In D. Reinking, M. McKenna, L. Labbo, & R. Kiefer (Eds.), *Handbook of Literacy and Technology,* Mahwah, NJ: Erlbaum.

Bopardikar, A., Sullivan, S. A., & Puntambekar, S. (2009, April). *Small group science talk in a design-based classroom: An exploratory study.* Proceedings of the annual meeting of the National Association for Research in Science Teaching, Garden Grove, CA.

Bryce, T. G. K. & MacMillan, K. (2009). Momentum and kinetic energy: Confusable concepts in secondary school physics. *Journal of Research in Science Teaching, 46*(7), 739-761.

Gerjets, P., Scheiter, K., Opfermann, M., Hesse, F. W., & Eysink, T. H. S. (2009). Learning with hypermedia: The influence of representational formats and different levels of learner control on performance and learning behavior. *Computers in Human Behavior, 25*, 360-370.

Lawson, R. & McDermott, L.C. (1987). Student understanding of the work-energy and impulse-momentum theorems. *American Journal of Physics, 55* (9), 811-817.

Mercer, N., & Littleton, K. (2007). *Dialogue and the development of children's thinking: A sociocultural approach.* London: Routledge.

National Research Council. (1996). *National science education standards.* Washington, DC: NAP.

Nussbaum, E. M. (2008). Collaborative discourse, argumentation, and learning: Preface and literature review. *Contemporary Educational Psychology, 33*, 345-359.

Olson, S., & Loucks-Horsley, S. (2000). *Inquiry and the national science education standards: A guide for teaching and learning.* Washington, DC: National Academies Press.

Puntambekar, S. (2006). *Learning from digital text in inquiry-based science classes: lessons learned in one program.* Proceedings of the 7th International Conference of the Learning Sciences (pp. 564-570).

Rouet, J-F., Potelle, H., & Goumi, A. (2005). The role of content representations in hypermedia learning: Effects of task and learner variables. In S.-O. Tergan & T. Keller (Eds.), *Knowledge and Information Visualization* (pp. 343-354). Berlin: Springer-Verlag.

Schvaneveldt, R. W. (1990). *Pathfinder associative networks.* Norwood, NJ: Ablex.

Schwarz, B. (2003). Collective reading of multiple texts in argumentative activities. *International Journal of Education Research, 39*, 133-151.

Smith, G. W., & Sullivan, S. A., & Puntambekar, S. (2009, June). *When to collaborate: Individual and group exploration of a hypertext environment within an inquiry science classroom.* Proceedings of the Computer Supported Collaborative Learning Conference, Rhodes, Greece.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Grounded theory, procedures and techniques.* Newbury Park, CA: Sage.

Varelas, M., & Pappas, C.C. (2006). Intertextuality in read-alouds of integrated science-literacy units in urban primary classrooms: Opportunities for the development of thought and language. *Cognition and Instruction, 24*(2), 211-259.

Wiley, J. & Bailey, J. (2006). Effects of collaboration and argumentation on learning from web pages. In A.M. O'Donnell, C. E. Hmelo-Silver & G. Erkens (Eds.) *Collaborative Learning, Reasoning, And Technology* (pp. 297-321). Hillsdale, NJ: Erlbaum.

Acknowledgements

CoMPASS research is supported by the National Science Foundation: IERI #0437660 & Institute of Education Sciences: CASL #R305A08050

Formulating WISE Learning Experiences

Vanessa Svihla, University of California, Berkeley, Graduate School of Education, Tolman Hall, Berkeley, CA, vsvihla@hotmail.com

Abstract: This study reports two cases of computer-supported activity drawn from intact classrooms implementing a Web-based Inquiry Science Environment (WISE) unit in the disciplinary context of global climate change. The two cases contrast balanced and uneven collaboration, and related learning activity. Interactions across scaffolded activity types are also contrasted, finding that narrative explanations provoked greater generativity as compared to other assessment formats. Analyzing the dyadic interactions from a Learning as a Member Perspective (LAMP) (Stevens, 2010), interactive visualizations emerged as much richer sources of learning opportunities, particularly when embedded in inquiry sequences. Design considerations are discussed.

Introduction

Technology enhanced learning commonly involves scaffolds intended to structure student learning through inquiry (Blumenfeld et al., 1991; Edelson, 2002; Quintana et al., 2004). Software scaffolds might guide students through phases of inquiry, such as exploring data, interacting with dynamic visualizations, and reporting findings, or help students track their progress. For instance, the Web-based Inquiry Science Environment (WISE) (Slotta & Linn, 2009) includes an inquiry map. Technology scaffolds may problematize the content (Reiser, 2004) or interrupt the deceptive clarity of a scientific model through activities that cause students to distinguish ideas (J. Chiu & Linn, in press). Pivotal cases, which provide an opportunity for students to restructure their ideas, may be scaffolded as predict-observe-explain sequences, and have been used in WISE units across various disciplinary topics as a means of making science accessible to students.

In a meta-analysis of studies on computer-supported learning, significant advantages were found for collaboration, particularly when completing a difficult task, with mixed-ability groups, and with single-sex groups (Lou, Abrami, & d'Apollonia, 2001). Technology may scaffold collaboration in numerous ways (e.g., Rummel, Spada, & Hauser, 2009)and this has been shown to be beneficial to developing integrated understanding (Bereiter & Scardamalia, 1989). The interactions may be logged when the technology also provides communication for physically distributed collaborators, but in face-to-face interactions between two participants using one computer, de facto collaboration cannot be assumed, even when deliberately scaffolded (Nussbaum et al., 2009). With co-located groups, interactions are reportedly very diverse (Volet, Summers, & Thurman, 2009). In such cases, "reciprocal interaction" (Rummel, et al., 2009, p. 72) may occur when there is symmetry in learning-related interactions (Dillenbourg, 1999). To evaluate the quality of collaboration, various dimensions have been identified: communication, information processing, coordination, interpersonal relationship, and motivation (Meier, Spada, & Rummel, 2007); problem solving contribution, frequency of participation (Barros & Verdejo, 2000); emotional state; collaboration behavior (Barron, 2003; Barron & Sears, 2002; M. M. Chiu, 2008); and sustaining mutual understanding, dialogue management, information pooling, reaching consensus, task division, time management, technical coordination, reciprocal interaction, and individual task orientation (Meier, et al., 2007). Recognizing the potential for students to differently frame learning when learning technologies are brought into a classroom, and particularly when they face a challenging task, Lantz-Andersson notes the need to analyze how students frame their interactions (2009). This study reports research on computer-supported dyadic activity occurring in middle school classrooms. Rather than assuming these are de facto collaborative learning interactions, two cases are presented to consider how students might (not) formulate the interaction as being a (collaborative) learning interaction. The *symmetry of contributions* across two dyads and across different activities is explored.

Global Climate Change

We designed a WISE unit called Global Climate Change, incorporating NetLogo visualizations (Wilensky & Stroup, 2000) representing the earth and atmosphere. The unit was designed according to the four main tenets of knowledge integration, making the content accessible, helping students learn from each other, making thinking visible, and promoting autonomous lifelong learning (Linn & Hsi, 2000). The knowledge integration pattern involves first eliciting ideas- both normative and non-normative, from the student's repertoire, then adding new ideas, then developing and applying criteria to distinguish ideas. The KI pattern foregrounds connections *between* ideas (Linn, 2006).

Methods

The participants of this study were two dyads drawn from two different intact middle school classrooms implementing the WISE Global Climate Change unit. The dyads were video recorded as they worked through the unit over seven days in 30-40 minute class periods. A high definition video camera was positioned behind each dyad, creating a record of the computer screen, computer usage, gestures that occurred between the two students, and direction the students faced. Thus, in most cases, students' faces were not recorded. Lapel microphones provided high quality audio recording of the conversations, even in the context of a crowded, noisy classroom. Finally, log files provided time-stamped sequences of actions on the computers. Videos were transcribed, including gestures visible between the students and actions taken with the computer, and all names were replaced with pseudonyms. Interaction analysis (Jordan & Henderson, 1995) focused on identifying moments in which students formulated the activity as being about learning, adopting a Learning as a Member Perspective (LAMP) (Stevens, 2010). Stevens argues that learning must "be shown to be a mutual concern of participants" (p. 84). This "endogenic" perspective on learning is useful for understanding the interactions that take place in a middle school classroom using "learning technology." Cases were not drawn with respect to their representativeness of the larger sample, but rather sought for their particularizability (Donmoyer, 1990); thus dyads were deliberately selected from a larger corpus based on individually-completed post-unit interviews. Cases were sought in which both students, and in which only one student could provide coherent explanations of what they had learned. A constant comparison method was used (Vogt, 2002), contrasting the two dyads, and contrasting the interactions across different activities in the unit (Table 1) to understand ways participants *might* or *might not* formulate their activity as collaborative learning.

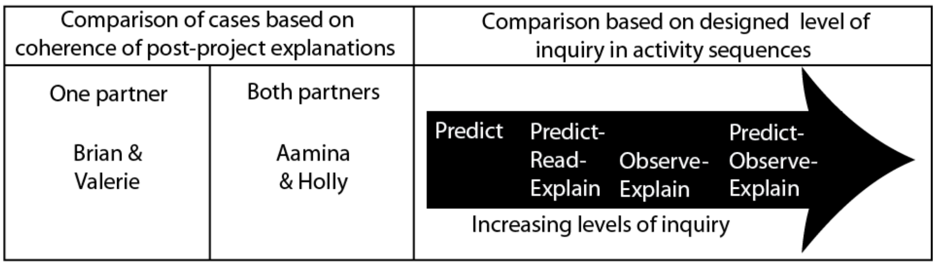


Figure 1. Dimensions Compared across Cases and across Activities.

Students first completed a pretest (Activity 1) in which they made various predictions including a narrative detailing the role of energy, then completed an activity designed to introduce a few key concepts (Activity 2). The remaining activities involved a sequence of interactions with NetLogo visualizations that let them experiment with variables to learn about global climate change (Table 1). Interactions with the first five activities are reported here.

Table 1: Five activities and their structure.

Activity	*1*	*2*	*3*	*4*	*5*
Predict	X	X	None	X	X
Observe	None	Text and figures	Solar radiation NetLogo	Albedo NetLogo	Atmosphere NetLogo
Explain	None	X	X	X	X

Results

Brian and Valerie: Uneven Collaboration

In their post-unit interviews, Brian provided a coherent explanation, including mechanisms and evidence, for global climate change, whereas Valerie struggled to explain even simple relationships from the unit, leading to a question of the origins of these two partners' very different understanding. For this case, gender roles may seem particularly salient. In transcript 1, for instance, as Brian and Valerie write a story to explain how the earth is warmed by energy, Valerie quickly abandons an accurate idea, then prompts Brian to provide a correct answer, which is arguably less accurate than her answer, then she affirms his answer. Although this interaction is not rich in terms of learning, it does lead Brian to hesitate and break ("and it moves? (.) How does it move?") from the routine process observed when answering more traditional multiple choice assessments, in which he quickly identifies the an answer and is confident on what is expected in terms of explaining his answer ("Yeah.

((indicating an answer)) And then say, like, that is the one we chose because, one day, means that global, global climate is heating up. It would have to be a long period of time.”).

Transcript 1

B: ((reading)) ‘Write a story to explain to Gwen how the earth is warmed by energy. Be sure to include: Where energy comes from::, How energy moves, Where energy goes, How energy changes/transforms.’ °okay ° So ener:: energy (0.9) ((beginning to type as he speaks)) comes (.) from (2.1) the (1.8) sun (2.2) and it moves (1.2) How does it move?

V: ((shifting computer towards herself)) It moves (.) in waves ((looks at B for confirmation, then very quickly shifting)) Just kidding! ((covers face, both laugh)). Um, you should know this.

B: Well, you should also.

V: I already did, I (inaudible) waves. ((she pokes him in the arm, then smiles))

B: I don't. (2.4) Energy comes from the sun and it moves (1.5) in rays.

However omni-relevant gender may be to these interactions, we can also observe other, perhaps more subtle, differences in how they interact as they move across the unit, engaging with inquiry activities that elicit their ideas, add new ideas, prompt for reflection and criteria-generation, or distinguishing ideas. Across these interactions, the roles the two students adopt vary, as do the interactions they have in relation to differing types of assessment steps and across sets of inquiry activities. For instance, Brian vacillates between two primary roles, teacher and learner. However, much of Brian’s “teaching” is not framed by an inquiry perspective. We can see various instances in which Brian adopts a teacher-as-task-master role, prompting Valerie to read aloud (“You read that part”), trying to get her back on task (“So some of 'em get deflected and some go in the earth and then -- Valerie!” and “So the purple ones -- Valerie! -- the purple ones are radiation”), giving directives related to being a good student (“Let’s listen to her ((the teacher))”), and correcting her spelling and pronunciation. For instance, as she is sounding out *infrared,* he says *inferred* and thereafter they talk about inferred energy. Some of Brian’s prompting of Valerie is about learning (“Don’t just guess”).Valerie sometimes formulates her role matched to Brian’s, as a task-completer, checking a box as instructed, typing what Brian says, and taking his directives. However, there is also evidence for Valerie formulating the situation to be about learning. When she asks a question (Transcript 2), Brian is busily engaged with his own learning and does not acknowledge her question (meaning he sometimes switches to a learner role). They attend to different aspects of the NetLogo visualization, such that although they are both in learner-roles, their learning cannot be considered collaborative.

Transcript 2

B: And then some go in the earth and create those little things or some (.) and then they go out in heat rays (3.9) Whoa. (1.9) Look how fast they’re going! ((Nudges V, they both laugh briefly. They click *watch a sunray* and follow its progress))

V: ((pointing to the screen)) Uh-huh. Cool. ((inaudible)) could never go that fast.

B: (16.9) Watch a sunray. Oh so those are the sunrays. ((pointing to sunrays)) That one went in but now it's ((his finger traces an expected path with the energy leaving the earth but the energy stays in the earth)) But now?

V: [What about that green line? ((pointing to land surface))]

B: [Now] Now it's doing conduction and now it went out as a heat ray. Okay.

As Brian and Valerie work on the albedo NetLogo, which is more inquiry-oriented than the previous activities, their interaction shifts as they both adopt learner roles (Transcript 3). While Brian remains unsure of what type of energy is being represented, Valerie simply describes the shapes (e.g., purple triangles) but in doing so, recognizes an important relationship that Brian had missed, that the “purple triangles” (infrared radiation) become sparse when there is high albedo. In this case, attending to differing aspects of the visualization is productive, at least for Brian who integrates his own understanding with Valerie’s observation when answering an Explain prompt (Transcript 4).

Transcript 3

B: It looks like the energies are ((pointing at the screen)) or whatever that ((indicating yellow triangles)) the solar radiation goes down, then up.

V: Ummhmm ((agreement))

B: ((clicks *reset*, then clicks *farms* again)) Farms.

V: Or like when the yellow ones are going down the purple ones are going up ((pointing up towards ceiling then down)) and when the yellow ones are going up the purple ones are going down

B: Farms ((clicking *reset* then *farms* again)) Most of ‘em (1.4) aren't getting absorbed=

V: =Umhm.

B: Er:: well. °Depends° (((21.0) clicks *desert* then clicks *reset* them clicks *desert* again, V leans in, B clicks *watch a sunray* and it reflects, clicks *watch a sunray* and it is absorbed)) So this one, for desert, it's going down. (10.2) Just keeps on going down (8.1) Okay! (((19.1) clicks *reset* then clicks *ice*, then clicks *ice* repeatedly- he does this each time- click-click-click-click-click then *reset*s, then clicks *ice* again. Drags *slider* again making albedo first low then back to high)) Okay it's ((referring to temperature)) going down. Look at it go.

V: Wow. (12.9)

B: ((clicks *watch a sunray*, which reflects))

V: There's like, no more purple ones.

B: Oh, ye::ah.(1.7) Not a lot of 'em are getting absorbed (.) It's reflecting it.

Transcript 4

V: ((reading)) 'Which of the following environments had the most [ener.

B: [Oh! The purple things! Well it wasn't the ice. The ice barely had any!

Aamina and Holly: Balanced Collaboration

In their post-unit interviews, both Aamina and Holly provided coherent, elaborated explanations for global climate change. The interactions between Holly and Aamina are very different from those between Brian and Valerie. Holly and Aamina modify each other's ideas (transcript 5, in which they are answering the same question presented in transcript 1) and negotiate their understanding. This interaction appears much more collaborative, and more frequently seems to be about joint learning.

Transcript 5

A: How energy moves:::s, um, ((taps fingers on edge of computer)) How does it move? Or where does it move? Oh where energy goes. How does it move? Hmm?

H: It moves through radiation? ((looking at A)) (29.7)

A: ((typing)) And then where energy goes?

H: Through the earth's atmosphere.

A: ((typing)) It enters earth's atmosphere::re (.) as heat? ((types))

H: (22.1) Or visible light?

Although the computer use appears to be imbalanced, with Holly more commonly in control of the keyboard, an important difference between the two dyads was observed; whereas Brian assigned tasks to Valerie who generally completed them as instructed, when Aamina directed Holly, she sometimes negotiated the task, or deferred it. For instance, when they begin the activity on albedo, the girls skip past the Observe step, believing it to be a repeat of the earlier NetLogo. When they reach the Explain prompts, Holly realizes they have missed something ("Let's go back"). However, when they reach the NetLogo, Aamina persists in thinking "We already did this," until Holly shows her the NetLogo (Transcript 6).

Transcript 6

A: We already did this. ((reading)) 'What happens to global temperature when you select an environment.'

H: ((scrolling down to the NetLogo, skipping directions))

A: Oh! So you can! ((Pointing at environments)) Let's do ocean.

H: Let's see. ((clicks *ocean*))

A: Let's hit run and then ocean.

H: ((clicks *run*))

A: Ocean.

H: ((clicks *ocean* again then *slows* model speed down))

A: Try to hit one of these ((pointing to environments)) I don't know what it's on right now.

As they continue to work with the NetLogo, Aamina offers directives which Holly accepts or defers. Aamina makes no attempt to take the computer control from her, but occasionally expresses uncertainty (Transcript 6, "I don't know what it's on right now.") in response to Holly's actions. Holly is less vocal than Aamina, but it is the negotiation of the interactions that makes this dyad seem more collaborative. There is evidence that the girls formulate this activity as a learning activity. When Aamina prompts Holly to return to the Explain step (Transcript 7, "So the global temperature decreases? Ok. I think we can go back to 4.4"), Holly acknowledges the suggestion but instead begins a new investigation (Transcript 7, "Yeah. So forest ((clicks *forest*)) Let's reset it and then watch it on forest"), indicating that she was still trying to understand the NetLogo.

There is also evidence that that although they, like Brian and Valerie, attend to different things in the NetLogo, they achieve coordination. For instance, Aamina notices that the NetLogo provides the percent of light each environment reflects (Transcript 7, "Oh! It tells you the albedo up here") while Holly is trying to determine this same information through the *watch a sunray* function. Once this information is shared, both girls move through the environments, sometimes with Aamina prompting, sometimes with Holly clicking. As they do so, they achieve a joint understanding and together conclude correctly that ocean has the lowest albedo.

<u>Transcript 7</u>

H: So, desert. ((clicking *desert*)) Desert. ((clicks *watch a sunray*)) and then ((pointing at screen)) so ice was, like, whitish.
A: Uhhh:::
H: ((clicks *farms*))=
A: =ummhmm=
H: =farms ((clicks *watch a sunray*)) (5.3)
A: So farms doesn't do that much.
H: Farms doesn't do it so we just have to check that one ((indicating and clicking *forest* then clicks *watch a sunray*))
A: Actually I think forest does it less than farms. And the ocean?
H: ((clicks *ocean* several more times, then clicks *forest* then *ocean*))
A: So the global temperature decreases? Ok. I think we can go back to four point four ((the explanation step that prompted Holly to return to the observe step))
H: Yeah. So forest ((clicks *forest*)) Let's reset it and then watch it on forest ((clicks *reset* then clicks *forest* several times, clicks *reset* several times, then clicks *forest* several times then clicks *watch a sunray*))
A: Oh! It tells you the albedo up here ((pointing)) for each one. That's forest, now try ocean.
H: ((clicks *ocean*))
A: Five percent ((still pointing, but moves finger over to *farms*))
H: ((clicks *farms*))
A: Farms 40 percent ((moves finger to *desert*))
H: ((clicks *desert*))
A: Desert's seventy-five percent
H: ((clicks *ice*))
A: Ninety-five percent=
H: =So they're=
A: =so the least has=
H: =ocean=
A: =yeah.

Representing Learning and Collaboration

Models of the dyadic interactions were constructed for the three activities containing NetLogo visualizations (Figure 3). To approximate the symmetry of participation, conversational moves and actions taken (e.g., clicking *ocean*) were covered by grey rectangles, approximating the relative length of turn (and with one members' utterances mirrored). Utterance length in seconds could have been used, but neither method is without flaws; as such the resultant representations are low fidelity models that allow for quick visual check for symmetry. Thus, time progresses from top to bottom, allowing the viewer to quickly judge symmetry, and reducing the likelihood of perceiving one member as positive and one as negative.

Overlain on this, are vertical bars representing computer control; when two bars of equal saturation are present, the computer was being shared, but when one of the vertical bars is of lesser saturation, the computer control was contested. Boxes indicate when students were interacting with a NetLogo visualization, and the sequences of activity are aligned by onset of the Observe phase of activity.

Finally, at the outer edges of the bounding boxes, are blebs indicating points in the transcript in which the students appear to be formulating the interaction as a *learning interaction.* For example, the interactions in transcripts two, three, five and seven were coded as cases in which the participants formulated their interaction as a learning interaction. Not all co-occurring learning interactions are collaborative because in some cases, particularly with Valerie and Brian, the questions they pose or seek to answer are not in a shared problem space. While the accuracy may be low, this approach provides a simple way to visualize potentially interacting dimensions as they unfold over a sequence of activity.

These representations (Figure 3) highlight both the differences between the two dyads, as well as differences across the steps. Most of Valerie's control of the computer is during explanation phases, and a much greater percentage of her interactions with the computer occur as contested, either because she is carrying out Brian's instructions, or because she is struggling to keep/gain control. Although some of the time Holly is in

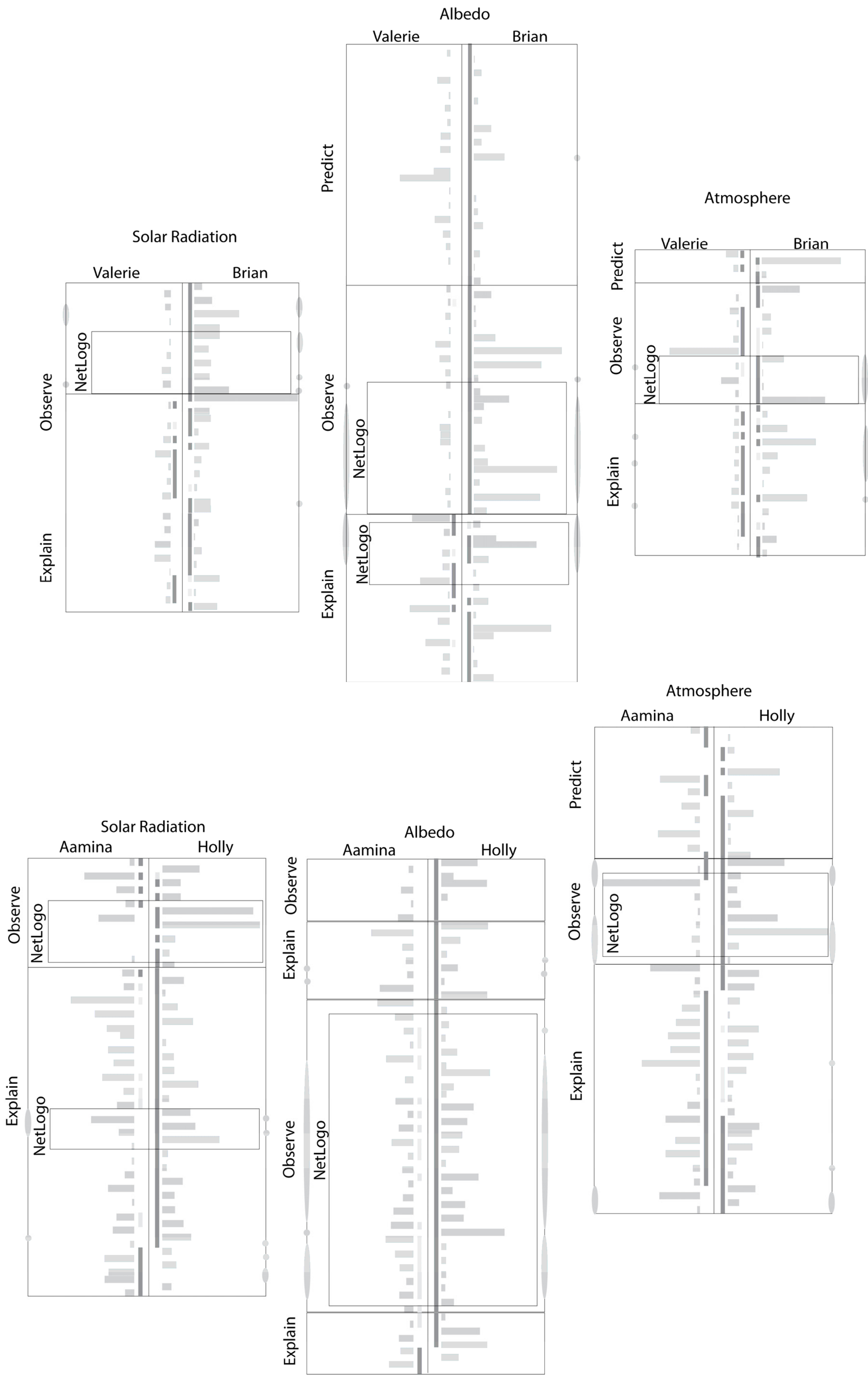


Figure 3. Representations of the Interactions across Predict-Observe-Explain Sequences.

control of the computer (punctuating their exploration of the albedo NetLogo, for instance) this is moderated by her willingness to resist, to renegotiate with her partner. It is also clear that Brian's turns are longer than Valerie's, whereas greater symmetry is seen in the turns taken by Aamina and Holly. Though Aamina is more vocal than Holly, because Holly negotiates computer control, she also contributes, making the interaction more symmetrical.

In both cases, the participants tend to formulate the interactions as learning while using the NetLogo visualizations, particularly when they are working with the more complex albedo and atmosphere NetLogos that are scaffolded by a series of predictions and explanations. In the case of Brian and Valerie, though they concurrently formulate the activity as learning, they do so in an asymmetrical manner, such that only Brian pursues the questions he poses. For Holly and Aamina, though they attend to different aspects of the NetLogo, they jointly pursue questions, resulting in interactions that can be considered to be both collaborative and learning.

Discussion and Implications

As has been noted elsewhere (Volet, et al., 2009), the interactions observed were diverse. In examining symmetry (Dillenbourg, 1999), or "reciprocal interaction" (Rummel, et al., 2009, p. 72) many dimensions provided relevant information for understanding the collaboration. In both cases, computer control is asymmetrical, as are verbal contributions, yet only in the case of Brian and Valerie was the learning distinctly asymmetrical. By adopting a Learning as a Member Perspective (LAMP) (Stevens, 2010), the asymmetry is understood to be consequential for Brian and Valerie; even when both formulated the activity as learning, they held different learning goals, and with Brian both controlling the computer and dominating the conversation, Valerie rarely pursued her questions, and her turns with the computer tended to occur during explanation phases, but functionally she did not explain her ideas; rather, she typed Brian's explanations. Given that teachers commonly encourage students to take turns "driving" the computer, it would be useful to provide guidelines to ensure that both students interact with complex simulations. A simple-to-implement solution might be to include an even number of steps between those containing NetLogo visualizations, though this would not likely have helped in the case of Brian and Valerie. Perhaps, in this case, the findings that mixed-gender groupings tend to be less beneficial is relevant (Lou, et al., 2001). Scaffolding students through a jigsaw activity in which partner is responsible for information that must then be integrated might support better collaborative learning as well, but further work is needed to understand whether such activity would actually interrupt unbalanced collaboration as seen with Valerie and Brian.

In addition to differences across the two dyads, the LAMP framing revealed differences in interactions by the type of prompt and level of inquiry. The narrative assessments are associated with more generative activity compared to the more traditional multiple choice plus explain prompts, the former of which tended to result in more attempts at negotiation, the latter of which were routinized and did not, in these particular instances, provoke discussion. Students have schema for answering such questions, and even when selecting incorrect answers and providing shallow explanations for their answers, students appeared confident in their responses.

The NetLogos that were scaffolded with predictions and observations tended to be associated with learning. While the initial questions students explored tended to be about clarifying the models, as in the first NetLogo, the explanation prompts encouraged the students to revisit the NetLogo visualizations to explore further. What emerges is that the more complex tasks tended to be when students formulated the activity as learning, and also that Holly, Aamina, and Brian all benefited from the observations of their partners, which relates to why previous research has noted an advantage to collaborative learning when completing a difficult task (Lou, et al., 2001).

References

Barron, B. (2003). When Smart Groups Fail. *Journal of the Learning Sciences, 12*(3), 307-359.

Barron, B., & Sears, D. (2002). Advancing understanding of learning in interaction: How ways of participating can influence joint performance and individual learning. *Computer support for collaborative learning: Foundations for a CSCL community*, 593-594.

Barros, B., & Verdejo, M. (2000). Analysing student interaction processes in order to improve collaboration. The DEGREE approach. *International Journal of Artificial Intelligence in Education, 11*(3), 221-241.

Bereiter, C., & Scardamalia, M. (1989). Intentional learning as a goal of instruction. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 361-392). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Blumenfeld, P. C., Soloway, E., Marx, R. W., Krajcik, J. S., Guzdial, M., & Palincsar, A. S. (1991). Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational Psychologist, 26*(3), 369-398.

Chiu, J., & Linn, M. C. (in press). The Role of Self-Monitoring in Learning Chemistry with Dynamic Visualization. In A. Zohar & Y. J. Dori (Eds.), *Metacognition and Science Education: Trends in Current Research*. London, UK: Springer-Verlag.
Chiu, M. M. (2008). Flowing Toward Correct Contributions During Group Problem Solving: A Statistical Discourse Analysis. [Article]. *Journal of the Learning Sciences, 17*(3), 415-463.
Dillenbourg, P. (1999). What do you mean by collaborative learning. *Collaborative learning: Cognitive and computational approaches*, 1-19.
Donmoyer, R. (1990). Generalizability and the single-case study. *Qualitative inquiry in education: The continuing debate*, 175-200.
Edelson, D. (2002). Design Research: What We Learn When We Engage in Design. *Journal of the Learning Sciences, 11*(1), 105-121.
Jordan, B., & Henderson, A. (1995). Interaction Analysis: Foundations and Practice. *Journal of the Learning Sciences, 4*(1), 39 - 103.
Lantz-Andersson, A. (2009). The power of natural frameworks: Technology and the question of agency in CSCL settings. *International Journal of Computer-Supported Collaborative Learning, 4*(1), 93-107.
Linn, M. C. (2006). The knowledge integration perspective on learning and instruction. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 243-264). New York: Cambridge University Press.
Linn, M. C., & Hsi, S. (2000). *Computers, teachers, peers: Science learning partners*: Lawrence Erlbaum.
Lou, Y., Abrami, P. C., & d'Apollonia, S. (2001). Small Group and Individual Learning with Technology: A Meta-Analysis. *Review of Educational Research, 71*(3), 449-521.
Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning, 2*(1), 63-86.
Nussbaum, M., Alvarez, C., McFarlane, A., Gomez, F., Claro, S., & Radovic, D. (2009). Technology as small group face-to-face Collaborative Scaffolding. *Computers & Education, 52*(1), 147-153.
Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., et al. (2004). A Scaffolding Design Framework for Software to Support Science Inquiry. [Article]. *Journal of the Learning Sciences, 13*, 337-386.
Reiser, B. J. (2004). Scaffolding Complex Learning: The Mechanisms of Structuring and Problematizing Student Work. *Journal of the Learning Sciences, 13*(3), 273-304.
Rummel, N., Spada, H., & Hauser, S. (2009). Learning to collaborate while being scripted or by observing a model. *International Journal of Computer-Supported Collaborative Learning, 4*(1), 69-92.
Slotta, J., & Linn, M. C. (2009). *WISE Science: Web-Based Inquiry in the Classroom*. New York: Teachers College Press.
Stevens, R. (2010). Learning as a Members' Phenomenon: Toward an Ethnographically Adequate Science of Learning. *National Society for the Study of Education, 109*(1), 82-97.
Vogt, F. (2002). No ethnography without comparison: the methodological significance of comparison in ethnographic research. In G. Walford (Ed.), *Debates and developments in ethnographic methodology* (Vol. 6, pp. 23-42). Oxford: Elsevier.
Volet, S., Summers, M., & Thurman, J. (2009). High-level co-regulation in collaborative learning: How does it emerge and how is it sustained? *Learning and Instruction, 19*(2), 128-143.
Wilensky, U., & Stroup, W. M. (2000). *Embodied science learning: Students enacting complex dynamic phenomena with the HubNet architecture.* Paper presented at the Annual meeting of the American Educational Research Association. Seattle, WA.

Acknowledgments

The author would like to thank the members of the WISE research group, the Video-analysis Seminar, and reviewers whose comments strengthened this paper, and acknowledge the NSF funding for this work, (Cumulative Learning using Embedded Assessment Results, CLEAR, NSF # 0822388) which does not necessarily represent the views of the NSF. Thanks are also due to the teachers and students who participated in various ways.

Students Producing Thick Descriptions

Richard Alterman, Johann Ari Larusson, Computer Science, Brandeis University
Email: alterman@cs.brandeis.edu, johann@brandeis.edu

Abstract: The work students do to appropriate *talking frames* is a significant element of the coursework for any class. The talking frames provide a viewpoint from which the course material "appears" connected and cohesive, from which "sense" can be made. The actual sense that is made is an explanation. The talking frames are semiotic tools that mediate the in-class discussion led by the teacher and later mediate an online collaboration, *sans* the teacher, in a co-blogging community. Assigned readings and the teacher's lecture provide talking frames for explaining course material. In the co-blogging environment, the students practice at "speaking" with the talking frames, *appropriating* the talking frames by using them to mediate the collaborative production of *thick description* about phenomena relevant to the course content.

Introduction

It is the teacher's job to explain how the elements of the course material fit together. The teacher's explanations use *talking frames* that compose the semiotic of the targeted field. The *appropriation* (Baker et al, 1999; Koschmann, 2002) of talking frames by the students is a significant element of the coursework for any class. In-class discussion provides some opportunities for the students to learn to explain with these "official" frames, opportunities to explain in the "appropriate" manner. Unfortunately, class time is a limited resource and there is insufficient time for students to practice. Outside of class, there is a coordination problem that makes it difficult for students to practice with one another. Online discourse communities potentially reduce the coordination problem by enabling students to discuss course material even though they work at "explaining" at different times and places. With this scheme, the talking frames are semiotic tools that mediate the in-class discussion led by the teacher and later mediate an online collaboration, *sans* the teacher, in a discourse community, where the students work at appropriating the talking frames. At issue is the form of the discourse community and the content that is produced.

This paper presents a case study of undergraduate students in an Internet & Society course who co-blog throughout the semester. As they co-blog, the students decipher the "codes" that compose the semiotic of the course material, learning how to explain the relevant phenomena with the talking frames they are learning. The students' online work in the co-blogging community manifests itself as the collaborative production of *thick descriptions* (Geertz, 1973; Ryle, 1968).

Appropriation of Talking Frames

There is an explanatory framework – the talking frame – and there are explanations. The talking frame provides a viewpoint from which the course material "appears" connected and cohesive, from which "sense" can be made. The actual sense that is made is the explanation. Given a talking frame, students may or may not converge on the same explanation for a piece of phenomena in the target domain.

An example of how this works is the accident at the nuclear power plan at Three-mile Island in 1979 (Gamson, 1992). Nine years before Three-Mile Island there had been a similar kind of incident in Detroit, but at that time the "interpretive packages" available were frames like *faith in progress* or *one step back for every two steps forward*. By the time the incident at Three-Mile Island occurred there were competing frameworks – *public accountability, small is beautiful* – for explaining the exact same kind of event.

In a class like Internet & Society the students are both learning the course material and also appropriating the talking frames, which they use to generate their explanations. It is the second part, the appropriation of the talking frames that is the subject matter of this paper. The students practice with the talking frames by using them to generate explanations of the relevant phenomena. This learning process can be described as *appropriation* (Baker et al, 1999):

> Children are said to appropriate cultural objects (material and semiotic tools), when they learn from other members of the culture how those cultural objects are used, and what they are used to accomplish. Appropriation is not a process of rote-learning, in which the individual simply adopts the facts and assumptions of the culture. Children appropriate these objects by participating in their use with more expert members of that culture. Learning results from the child's own experiences and practice with the object under the guidance of an expert. Thus, it is not a matter of information transmission from the expert to the novice, but of the novice "making this tool his own" (Leontjev, 1981).

The goal is for the students to *appropriate* the semiotic of the teacher for discussing the course material. The talking frames the teacher uses in class are the semiotic tools that mediate the in-class discussion. In the blogosphere, the students generate explanations using these frames *sans* the teacher. As the students contribute to the blogosphere, the talking frame, a semiotic tool, is appropriated. The trajectory is from an interpsychological plane with the instructor, to one without the instructor, to an intra psychological plane (Vygotsky, 1978). In each case, the talking frames are the semiotic tools that mediate the activity.

Initially the line of arguments, the explanations provided by the teacher are thin, skeletal. The students need to work at making them more bushy, exploring their meaning, seeing how they interact with other ideas and explanations, filling out the details, making it substantial. In the blogosphere, students are deciphering codes and producing *thick descriptions* (Ryle, 1968; Geertz, 1973) of the course content. They are producing "piled up structures of inferences and implication" (Geertz, 1973). Ideas and concepts are poked, prodded, and played with. The goal here is not convergence. The students need to play with the talking frame. Use it. Analyze it. Run it, and see how it works. The externalization of descriptions, reflections, explanations, analyses, and arguments makes them accessible to other students, *sedimenting* (Stahl, 2006) the collaborative sense that is made with the talking frames available to the class. The talking frames become a part of the parlance of the students, but agreement on their exact usage is not required. Because the appropriation of the talking frame requires play, the more the students produce explanations and descriptions with a given talking frame, the thicker will be the set of descriptions, the more it is a part of the assumed background knowledge of the students.

To summarize this discussion:

1. Assigned readings and the teacher's lecture provide talking frames for explaining course material.
2. In a co-blogging community, the students practice at "speaking" with the talking frames.
3. In a co-blogging community, students appropriate the talking frames by using them to mediate the production of thick descriptions about phenomena relevant to the course content.
4. Through the use and application of the talking frames, the students enrich their common background knowledge, collectively producing better, richer, more nuanced and textured descriptions, explanations, analyses, and arguments.

Co-blogging as Practice with Semiotic

A *discourse community* is where students question, criticize, explore, negotiate meanings, share expertise, constructing and developing new understanding and a "common mind and voice" (Brown et al,1993; Wertsch, 1991). Participating in an online discourse community is a social activity. Students can collaborate without "meeting" at the same place or at the same time, thus increasing the opportunities for fruitful discussion.

Within a discourse community, the contributions of the students are published and broadcast to the rest of the class, emerging in an open space, giving students exposure to multiple viewpoints and perspectives (Stahl, 2001; Suthers, 2005). Knowledge communities, arguing to learn, discussion forums, and co-blogging environments are all collaborative activities that to a greater or lesser degree function as discourse communities. The students can explain to one another, argue, negotiate, and reflect, which are all activities that positively impact learning (de Vries et al, 2002; Andriessen et al, 2003; Salmon, 2002; Scardamalia & Bereiter, 1996).

In the study presented in this paper, the students participate in a discourse community: they co-blog throughout the semester. In the co-blogging community, each student has a *blog*. The blog is composed of multiple *posts* written by the blog owner. Students can read each other's blog posts and *comment* on them. Some features of the student co-blogging activity are:

1. **Volume:** Posts link to other material in the blogosphere or refer to material introduced in other venues like the lecture, but they are also self-contained. Each post is relatively long: the average post length during the semester was 277 words and the median 248 words.
2. **Open format:** Students have freedom to self-select the mode of their contribution, e.g., reflective, argumentative, analytic, or recapitulative (Du & Wagner, 2005). Less scaffolding reduces the time it takes to learn and to use the co-blogging application.
3. **Emergence of identity:** Each student has her own blog, and she has full control over its content, thus establishing personal and intellectual ownership of her work (Ferdig & Trammel, 2004). Students develop an individual style and voice.
4. **Convergence is not necessary:** Students do not converge on any preferred viewpoint. Each post, with its comments, is a self-contained conversation that explores some of the implications of the initiating post. The blogosphere supports a more textured view of complicated ideas without requiring agreement.
5. **Overhead:** Because co-blogging is Web 2.0 technology, the "buy-in" for students is fairly cheap (Glogoff, 2005; Duffy 2008). Because co-blogging has relatively less scaffolding than other forms of discourse community, the overhead in learning to use the co-blogging environment is low: a typical student takes one in-class lab to learn to co-blog.

Students in a co-blogging environment are free to explore talking frames, deciphering their meaning and usage as they explain relevant phenomena. Contributions can vary among alternate modes of discursion, including reflection, argument, analysis, (re-)articulation, and story telling. The openness, volume, and lack of convergence are all contributing factors to the emergence of the co-blogging environment as a space to practice with the semiotic of the course.

Case Study

In the Internet & Society course taught in Fall 2008, 25 students collaboratively blogged throughout the semester. The course explored the impact of the Internet on society. Topics for the course included the Internet revolution, identity, information versus knowledge, technology and social inclusion, copyright and patent law, and democracy and the Internet. An important part of the class was for the students to integrate what they learned with their general knowledge about the everyday world they live in. The students were undergraduates from a variety of disciplines. There were 3 science majors and 1 science minor in the class. There were 12 students majoring in the social sciences and 8 minoring in the social sciences. The remainder of the class was either in the humanities or fine arts. There were 8 females and 17 males.

Lectures were presented using slides that summarized the key points of the lecture. At the beginning of each lecture, hard copies of the slides were handed out to support student note taking. PDF versions of the slides were downloadable from the class website. The lecture slides were used as a basis for identifying the important course topics and themes.

The co-blogging environment has been developed over a number of years in several different courses (Larusson & Alterman, 2009). The students were required to blog at the pace of one post per lecture: there were two lectures per week. A typical post was 2 paragraphs in length. The students were also required to read and comment on other contributions to the blogosphere. The co-blogging work of each student counted for 35% of his or her grade.

All of the students' online work was automatically recorded in a transcript, which enabled both quantitative and qualitative analyses. The transcripts can be treated as an event log file and accessed using database queries. Additional tools enable a large variety of alternate analysis methods, including discourse and ethnographic.

At the end of the semester a survey was distributed; questions were on a 6-point Likert scale (from 1, not useful, to 6, very useful); the survey also included open-ended questions. Student assessment of the co-blogging activity was positive. When the students were asked to rate the value of their online co-blogging work as a means of giving them first-hand experience with online collaborative learning, the average response was 5.6. In response to the question of whether the students felt the co-blogging community was useful, the average response was 5.3. When queried about the usefulness of the blogosphere for writing papers, the average response was 4.5. When asked as a yes/no question whether re-reading and reusing the blogging text helped the students write their papers, 67% answered in the affirmative.

The Blogosphere

The lecture and discussion in the class is a collaboration between the teacher and the students to understand the material. But the collaboration is of a particular kind. The collaboration is asymmetric. There is an authority, the teacher, who is the presenter – the "guardian" of the "official story" on the course material. The teacher employs talking frames that reveal important relationships within the material. The texts of the course are "coded" in the semiotic of the field, couched in forms and terms that compose an "authentic" discourse. When the discussion moves to the blogosphere, the conditions of the collaboration change significantly. The blogosphere is a student-owned space (Oravec, 2002). The students must "appropriate" the teacher led dialogue. The students learn to "talk" in their own "voices" about the material: they are practicing and learning to talk and reason about the material, firming up their "grasp" of the material, collaboratively thickening the initial descriptions and explanations, embellishing and adding to the skeletal structure that was first presented in an assigned reading or during lecture.

Contributions to the blogosphere refer to and re-articulate course topics and themes, developing the associated talking frames. The contributions explore the semiotic of the course, deciphering codes as they layer descriptions of the relevant phenomena.

Contributions also refer to the common background of the students, either as co-participants in the class (e.g. their common experience of reading an assigned text) or as undergraduates at the same institution (e.g., using Google to help do their schoolwork). The cases, examples, issues, analyses, and arguments that are drawn from the common, shared, and individual backgrounds of the students further ground each contribution and the conversation it engenders.

Contributions to the blogosphere have one or more talking points, issues that are addressed in the post or comment. The talking points function to coordinate student effort to expand on, and refine, the individual and collective conception of codes and their application. The bulk of a post is grounded in examples and cases

(Kolodner, 2006), which may be drawn from personal experience or transcribed from other course-related activities like lecture or reading the text. The *mode* of a contribution can vary: an individual post can alternate between discursive modes of argument, (re-)articulation, reflection, analysis, and story telling.

The opening text of a blog post is where a student directly or indirectly connects his or her commentary to what he or she assumes is common knowledge among the students in the class (see Figure 1). Students explore the semiotic of the class by referring to and quoting the text, explicitly referring to examples, arguments, and distinctions made in the text (1). Other semiotic relevant activities they introduce into the ongoing discourse are in-class discussions (2), and the presentation slides of the instructor (3). Students ground their contributions in experiences that are likely to be common amongst the students like other activity in the blogosphere (4), internships (5), googling (6), and studying abroad. Frequently posts begin with multiple kinds of links, mixing together references to assigned readings and class discussions (semiotic) with common experiences (ground) like pop culture (7) and previously introduced cases (8).

1. Information vs. knowledge is an important distinction presented in The Social Life of Information. I have to agree that information on its own is not good enough, not complete enough to qualify as knowledge. I see knowledge as ...
2. Talking in class about limits of information on the internet made me think that we cant expect internet to solve all ours problems at once. The cant expect that Amazon will buy us book automatically catered to our preference. We cant expect...
3. I've been thinking a bit about the limits of information, especially the second to last question asked in the slides: "What kind of critical commentary does this lead you to make about the huge amounts of information that are being generated on the Internet?"
4. After doing some reading and viewing some blogs, I began to feel that the amount of information that is offered to internet users is having a major effect on the world and the type of news that is being presented. One of the blogs commented on the question that was presented in one of the class slides: "What kind of critical commentary does this lead you to make about the huge amounts of information that are being generated on the Internet?"
5. Each summer millions of students participate in summer internships in a diverse range of industries. Students spend countless hours researching the fields they are interested in. Students acquire information about the technical skills necessary as well as the soft skills they will need to apply in the workplace. This is the accumulation of information. Last summer I engaged in my first investment banking internship …
6. What I found to be very useful whenever I have problems with my computer is to google it...
7. On page 21 of their book, John Seely Brown and Paul Daguid assert that companies that use the prefix "info" multiply their market values. Considering how so far about 4 people have read my blog, I'm expecting 12-16 comments for this one. I'm going to continue with the pop-culture themes I've used in my other blogs, because, well, that's how I think....
8. On the other hand, like Nadine mentioned how her friend and her dad works from home. I know that I can't concentrate when I am in my room... but when I am able to concentrate I can do the work anywhere. I get distracted when I am in my room, then again when I have a test or something it does not matter to me where I study because I stay focus.

Figure 1. Excerpts from Posts.

In the blogosphere, the range of discussion is broad and bushy with multiple viewpoints, and conversations, emerging. Each post is the root of a new "conversation" tree. Each conversation is self-contained, encapsulated, but it can also link to other blogs and conversations within the blogosphere. Discussions develop as smaller chunks of interaction. The opening of a comment links to the initiating post and/or earlier comments on the post. Each comment on a post links to at least one of the talking points in the initiating post; less often they respond to an earlier comment. Sometimes a contribution to a conversation will link to a different conversation. For each theme or topic several conversation trees may develop. Comments can agree and expiate on a contribution made in the initiating post or negotiate over its conclusion, providing alternate arguments, examples or cases to consider – all of which contribute to the collaborative production of thicker descriptions.

Co-blogging on the Frame *Working Home Alone*

A tremendous amount of information is produced each day on the Internet. Can this information solve/change everything? Can agents/bots help us to manage all the information? What is the difference between information and knowledge? Does access to information mean that people will be able to *work home alone* and that they no longer need to work in the office? How much does collocation matter? Issues like these are explored in The Social Life of Information. Although they have a balanced view, Brown & Duguid (2002) are intent on showing the limits of information. Their exploration of these issues provides many motivating examples, presenting alternate viewpoints, and arguments.

One issue in The social life of information concerns the pros, cons, and conditions of *working home alone*. This is a theme that has several topics associated with it. Does the information available on the Internet free people to work home alone? Is collocation at the office still necessary? Working home alone is a talking frame that can be used to explain a number of things regarding the limitations of information, the relation of information to knowledge, and so on.

How does the co-blogging activity of the students thicken the initial talking frame for working home alone? The students re-articulate, clarify, expand upon, and contextualize the semiotic introduced by the lecture and assigned texts. They explore, expand, contest, and re-explain the major arguments. They provide cases and examples that further deepen the discussion, grounding it in the individual and common experiences of the students. In the blogosphere, the students collaboratively develop a thicker description of the course material: a more nuanced and textured position on when and how the talking frame applies to the relevant phenomena and what it means and its limitations.

There were 6 posts and 10 comments on posts with regard to the topic of *working home alone* (see Table 1). There were a total of 3088 words produced collaboratively in the blogosphere on this theme. The average post length was 341 words; the longest post was 507 words and the shortest 113. The average comment length was 104 words. Taken together the posts weigh the pluses and minuses of working home alone versus working with others in the same location. The posts and comments develop descriptions, explanations, and arguments of the conditions under which working home alone will be effective and conditions where collocation matter**.**

The first post (P1 - 314 words) on this topic begins with a quote of The Social Life of Information text. The post mentions some advantages of being able to work home alone – e.g., savings on commute time. It also references the discussion in class of an example presented in the text: the difficulties of fixing technical problems when working home alone. The key talking point analyzes isolation as a potential problem of working home alone, concluding that working home alone should not "eradicate the office place." The *one comment* (c1) on this post links to the theme of isolation, grounding the commentary in the common student experience of "attending classes and living with fellow students." In the space of this one post and one comment, links are established to the text and class discussion and the collective description of these topics is re-articulated and thickened with the addition of new talking points on isolation and the common experience of all students in the class.

Table 1: Summary of collaborative thickening of talking frame working home alone.

	LINKS (SEMIOTIC & GROUND)	THICKENING ELEMENT(S)
P1	Text (quote, issue)	Re-articulation; analysis (isolation)
c1	Link to TP in post	Example (college experience)
P2	Text; class discussion	Re-articulation; 2 new cases (Friend's brother; Father); position
c1	Responsive to post	Example (home-business); analysis (discipline)
c2	1st comment (discipline); post	New example (studying library or room)
c3	Discussion (text; class); c1&2;P1	Re-articulate; common exp. (Brandies); analysis (self-discipline)
P3	Other co-blogging activity	Re-articulation; analysis (time); reflection (studying library)
c1	Text (quote, argument); P1&3	Re-articulation; argument
c2	Post (argument)	Example (friends); analysis (social aspect)
c3	1st comment (quote, arg.); P3	Re-explaining post; Re-articulation; position
P4	Common exp. (internship)	2 new Examples (internships); analysis (collocation)
c1	Post (example)	Re-articulation
P5	Text (argument; example)	Re-articulation; position
P6	Text (argument)	2 new cases; argument
c1	Post (argument; example)	Example (Online colleges)
c2	Post; c2 of P2; P3	Re-articulate; analysis(discipline; perseverance); position;example

The second post (P2 - 507 words) opens by linking to both the class discussion and book. The bulk of the second post presents two *cases* of people who successfully worked home alone, one is a friends' brother and the other is the poster's father. The third and last paragraphs of the post take a position that counterbalances (qualifies) the two positive cases she just discussed, concluding that working home alone can be very successful but "people need to be able to motivate themselves," the office "will not become obsolete any time soon," and the office place is still "useful for teaching newcomers the things they need to know." (The last quote refers to an argument made in the book.)

The *first comment* on this post talks about niche markets and home-based businesses: the example is part time work selling insurance through forums and chat rooms. The talking point picks up on the post's point about motivation: working home alone requires "discipline." The *second comment* links to the first comment on the discipline to work home alone: the author of the second comment goes to library rather than working in his room, and "totally" agrees with post that this does not mean people cannot get work done in their rooms. A *third comment* refers to argument about fixing technology when working home alone that was in the book, discussed in class, and re-articulated in the post. The student also acknowledges the point on self-discipline that was made in the two previous comments. His comment also addresses the talking point of studying in one's room, grounding his discussion in the cost of commuting to school to study: "Why go through all the hassle?" The conversation that emerges from post two is self-contained in the sense that the initial post reproduces enough context for other students to carry on a discussion. Old arguments are re-discussed. For the second time, the point on commuting is explored. Between the first conversation on working home alone and the second, the point on commute time has been explored and the following factors explicated: yes, working home saves commute time (post 1), but loneliness/isolation reduces productivity and collocation increases it (comment on post 1), but on the other hand, with self-discipline (comments 1 and 2 on post 2) and with the evidence provided by two examples of people working home alone (provided by post 2), why should a student go through the hassle of commuting (comment 3)?

The third post (P3 - 434 words) was written by the student who wrote the second comment on the second post, roughly two hours after writing his earlier comment. (Over the entire semester, roughly 35% of the reading events occurred while students were authoring blog posts.) The third post begins by linking to the previous blogging activity: "I know there are a few blogs about working home alone, but I still want to write one more just to let my thoughts out." It also specifically links to the second post: "On the other hand, like Nadine mentioned …" One talking point in this post concerns time management and the skill set of the person working home alone. Working home alone "all depends on the person." This author of the post reflects that he cannot concentrate when he is in his room (reflection); this implicitly refers to previous conversation.) In the second paragraph of this post, the student re-capitulates, in words of this student, the issue of dealing with technical problems when working home alone.

The *first comment* on the third post links to the discussion from an earlier post: it uses the word "isolation" which was originally used in the first post. The first comment also attempts to find neutral ground between the pros and cons of working home alone, repeating a line from the initiating post: "it all depends on the person." The *second comment* opens by linking to the phrase "it all depends on the person." Where the first comment picks up the theme (from the book and lecture) that working home alone misses the social aspect of office situations and the possibilities for learning, the second comment mentions friends who work "remotely" but would have been more effective in the office. The *third comment* is written by the author of the initiating post. It is an attempt to clarify his position: he was focusing on how it is possible to work from home, it is an option, but agrees with first comment on social networking – but then again people working home alone may have different priorities.

The fourth post (P4 - 248 words) provides two cases, personal examples of internships. In one case the interns collaborated online: 30 interns working in different places were able to interact weekly in an effective manner via a web seminar. In the other case, interns worked in an office. The student intern asked his supervisor whether he thought his job could be outsourced. The answer was no, because financial advising is based on trust, which depends on a face-to-face meeting. The one comment this post received says that there is "a level of employees who need to meet customers in person" but also predicted that "we will see lot more types of internships in the future online."

> The first four posts on the topic "working home alone" occur within a week of the lecture that discussed this topic. Half the comments were contributed to the blogosphere within that same time period. After a gap of over two weeks, two additional posts were made on this topic (these were made by students who had fallen behind on their blogging).

The fifth post (P5 - 113 words) is short. It refers to an example taken from the text that was also briefly discussed in class. It personalizes the discussion, declaring some people "just prefer to work home alone", explaining that if you can deal with problems that may arise (because you have the needed background), then working home alone may increase productivity and creativity. There were no comments on this post.

The sixth post (P6 - 432 words) is the last post on this topic. It is a lengthy post and it makes "the depends argument": it is written in a way that shows the student did not read the earlier discussion on the topic of working home alone. The post extensively discusses each of the student's parents, in one case working home alone was not appropriate and in the other case it worked. In case one, her mom is a project engineer who needs to be in contact with her group. In the case of her dad, he is a software engineer, who does a majority of his work at home. Despite the fact that the focus of discussion in class had moved on to another of the assigned books, this post attracted two comments. The *first comment* agrees with the points about the dad's job, and claims the post makes the perfect argument against online colleges, which is an example discussed elsewhere in the book. The *second comment* is made by a student, who earlier, within a week of the relevant lecture, had made a comment on the second post on this topic. The commenter says there is no right or wrong answer. It "boils down" to one's discipline (a recurring talking point) and perseverance to excel/succeed. The student's commentary is grounded in the example of student study venues: equal number of students who prefer to study at home versus study in library. This example was initially presented in the discussion of the second post: "Students ultimately choose their study spots according to their effectiveness and productivity at these places." This was also discussed in post 3.

Thick Descriptions

Figure 2 shows a tree map of student contributions to the blogosphere on topics related to the book The future of ideas. There are five major themes for the book. The theme "control" covers the largest area in the tree, which means there were more topics associated with this theme than any other theme. Since the theme "three layers" (which is in the upper right corner and mostly hidden) covers the least amount of space in the map so it had the fewest number of topics.

Each rectangle within a given area represents a topic. The brighter the color of the rectangle, the greater the number of contributions in the blogosphere on that particular topic. In terms of thickness these are topics that accrued the thickest set of descriptions. For example, the three brightest rectangles for the theme *new rules* are "versus old rules", "vinyl records and CD's", and "new ways of music production and distribution".

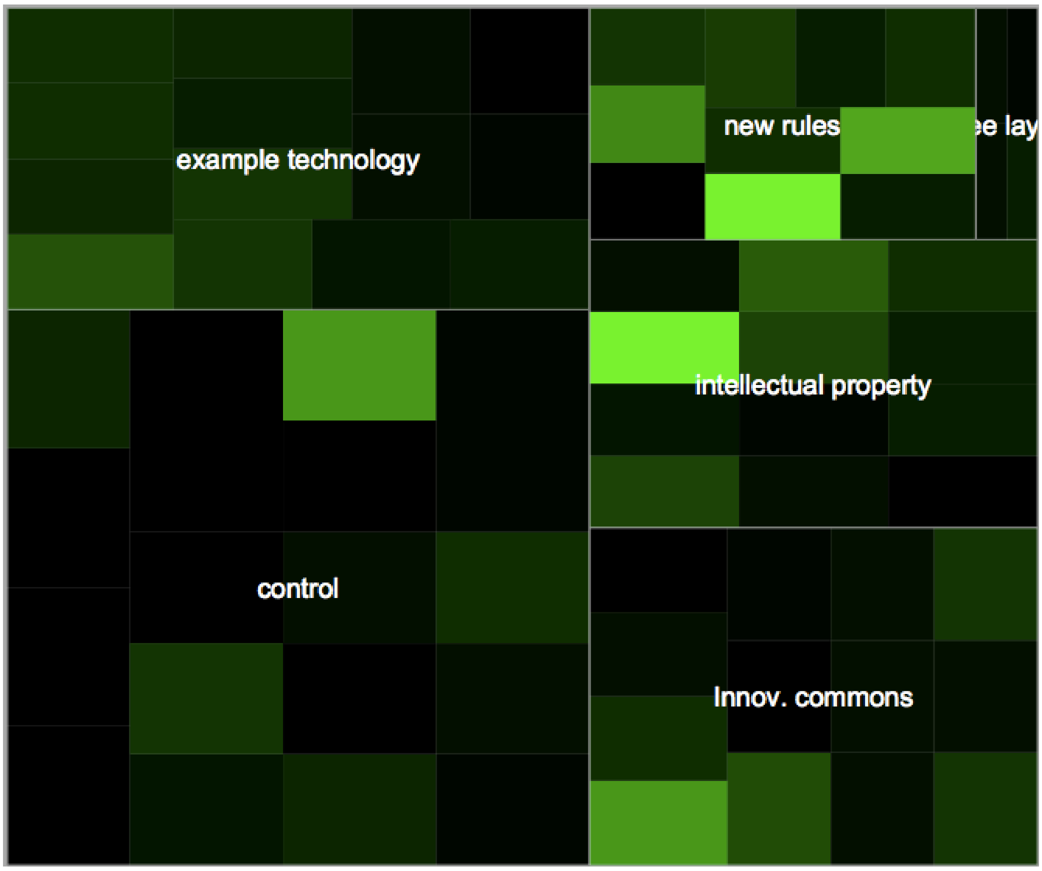


Figure 2. Tree Map of Contributions to Blogosphere on the Book The Future of Ideas.

It is not coincidental that the topics where the descriptions were thickest are topics that directly relate to the common interests and experiences of the students: the two brightest rectangles for the entire tree map are "copyright law" and "new rules of music production versus old rules". Why would students be interested in these topics? The students in the class grew up with the *new rules* of distribution being the norm. The topics for this theme have to do with copyright, distribution and access to multimedia content, and making money. It is easy to ground discussion in the students' shared interests in music and movies and their concerns about copyright infringement, making money, and being rewarded for creativity. They are the biggest consumers and the largest population of users practicing "civil disobedience" and "youthful rebellion" against the old rules.

Concluding Remarks

The exploration and appropriation of the *talking frames* is a significant element of the coursework for any class. It is the teacher's job to explain the course material, explaining how the elements of the course material fit together. The teacher's explanations use the relevant talking frames. The students *appropriate* (Baker et al, 1999; Koschmann, 2002) these talking frames by practicing at talking with them about the course material and other relevant phenomena in a co-blogging community. The trajectory is from an interpsychological plane with the instructor, to one without the instructor, to an intra psychological plane (Vygotsky, 1978). In each case, the talking frames are the semiotic tools that mediate the activity. The students' online work manifests itself as the collaborative production of a "many-layered sandwich" of *thick descriptions* (Geertz, 1973; Ryle, 1968). Ideas and concepts are poked, prodded, and played with using the talking frames. The talking frames become a part of the parlance of the class. The more the students produce explanations and descriptions with a given talking frame, the thicker will be the set of descriptions, the more it is a part of the assumed background knowledge of the students.

References

Andriessen, J., Baker, M. and Suthers, D. (2003) Editors, Arguing to Learn. Kluwer Academic Publishers.

Baker, M., Hansen, T., Joiner, R., and Traum, D. (1999). The role of grounding in collaborative learning tasks, In P. Dillenbourg (Ed.), Collaborative learning: cognitive and computational approaches. 31-63, Pergamon.

Brown, A. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of Learning Sciences*, *2*(2), 141-178.

Brown J. & Duguid, P. (2002). The social life of information. Harvard Business School Press.

de Vries, E., Lund, K., & Baker, M. (2002). Computer-mediated epistemic dialogue: Explanation and argumentation as vehicles for understanding scientific notions. *Journal of the Learning Sciences*, *11*(1), 63–103.

Du, H. S., & Wagner, C. (2005) Learning with Weblogs: An empirical investigation. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05)* (pp.7b). Washington, DC: IEEE Computer Society.

Duffy, P. (2008). Engaging the YouTube Google-eyed generation: Strategies for using Web 2.0 in teaching and learning. *The Electronic Journal of e-Learning, 6*(2), 119–130.

Ferdig, R. E., & Trammel, K. D. (2004) Content delivery in the "Blogosphere". *T.H.E. Journal Online: Technological Horizons in Education*, *31*(7), 12–16.

Gamson, W. (1992) Talking Politics. Cambridge University Press.

Geertz, C. (1973) Thick description: Toward an interpretive theory of culture. In The Interpretation of Culture, Basic Books.

Glogoff, S. (2005). Instructional blogging: Promoting interactivity, student-centered learning, and peer input. *Innovate: Journal of Online Education, 1*(5).

Kolodner, J. (2006). Case-based reasoning. In Swayer, R. (Editor), The Cambridge Handbook of the Learning Sciences.

Koschmann, T. (2002) Dewey's contribution to the foundations of CSCL research. *Proceedings CSCL,* 17-22.

Larusson, J.A., & Alterman, R. Wikis to support the "collaborative" part of collaborative learning. *Computer-Supported Collaborative Learning*, 4(4), 371-402, 2009.

Oravec, J. A. (2002). Bookmarking the world: Weblog applications in education. *Journal of Adolescent & Adult Literacy*, *45*(7), 616–612.

Ryle, G. (1968) The thinking of thoughts: What is 'Le Penseur' doing? *University Lectures,* no. 18.

Salmon, G. (2002). Mirror, mirror, on my screen … Exploring online reflections. *British Journal of Educational Technology*. 33(4), 379-391.

Scardamalia, M. and Bereiter, C. (1996) Computer Support for knowledge-building communities. In Koschmann, T. (editor), CSCL: Theory and practice of an emerging paradigm. LEA.

Stahl, G. (2001) *WEBGUIDE: Guiding Collaborative Learning on the Web with Perspectives.* Journal of Interactive Media in Education, 2001 (1)

Stahl. G. (2006) Group Cognition. MIT Press.

Suthers, D. (2005) Technology affordances for intersubjective learning: A thematic agenda for CSCL. *CSCL.*

Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Wertsch, J. (1991), Voices of the Mind: A Sociocultural Approach to Mediated Action, Harvard University Press.

Acknowledgments

Thank you to all four reviewers for their thoughtful comments on the submitted version of this paper.

Technology for Supporting Learners in Out-of-School Learning Environments

Tamara L. Clegg, University of Maryland, College Park, 2117A Hornbake, South Wing, College Park, MD 20742, tclegg@umd.edu
Christina M. Gardner, School of Interactive Computing, Georgia Institute of Technology, 85 Fifth Street, NW, Atlanta, GA 30308, Christina.Gardner@gatech.edu
Janet L. Kolodner, School of Interactive Computing, Georgia Institute of Technology, 85 Fifth Street, NW Atlanta, GA 30308, jlk@cc.gatech.edu

Abstract: In this paper, we present an evaluation of the role that software played in promoting collaborative scientific participation in a learning community driven by learners' interests and goals. In our analysis of an out-of-school learning environment we designed and implemented to promote scientific engagement, we found that the software was not always at the forefront in the learning environment. However, it played an essential role in providing an organized repository of the community's experiences and placed this information at learners' and facilitators' fingertips.

Introduction

We aim to help learners see the value of science in their everyday lives and come to see themselves as scientists. But many learners are turned off by science, finding it boring and irrelevant (Chinn & Malhotra, 2001). Many learners also face difficulties engaging in scientific inquiry. For example, learners often end investigations prematurely, forget the purpose of experimentation, and fail to recognize the importance of scientific situations (Gleason & Schauble, 1999; Quintana et al., 2004). To help learners overcome these difficulties and become interested and engaged in science, education researchers suggest we engage learners in the type of scientific participation scientists engage in. This type of scientific participation involves addressing real world problems that are interesting to learners and collaboratively explored with others who share their interests (Chinn & Malhotra, 2001; Crowley & Jacobs, 2002).

Our approach to addressing these needs has been to move outside of school and engage learners in science in contexts that are motivating and relevant to their lives. We have started with the context of cooking. We created a learning environment where participants engage in scientific practice by designing cooking experiments to learn about scientific phenomena underlying the dishes they make (e.g., what role do eggs play in brownies?). They then use their understanding of how these phenomena work in their dishes to perfect more complex dishes to suit their tastes. In this out-of-school context, we aim to promote the development of a community of learners that engages in science in contexts relevant to their lives. This community does not simply consist of learners, but also mentors or facilitators, older adults who share a love for cooking and value for scientific investigation with learners.

While cooking is a rich context for scientific exploration, there are many challenges to helping learners engage scientifically in contexts where learners are participating in order to pursue personal interests. It is often difficult to be prepared to help learners pursue their own interests and goals (Clegg & Kolodner, 2007). Furthermore, cooking activities are physically demanding. In the midst of busy, messy, and exciting projects, learners need help focusing on the relevant scientific aspects of their experiences (Gardner, Clegg, Williams, & Kolodner, 2006). We therefore designed software to scaffold learners' collaborative cooking experimentation in hopes that it would promote scientific engagement as they pursued their cooking goals. In this paper, we present an evaluation of the role that software actually played in promoting collaborative scientific participation in an active learning community driven by learners' interests and goals.

Background

The aspects of scientific practice that learners can best learn through cooking activities are design of investigations, interpretation of data, use of data as evidence, and explanation. Such practice includes, but is not limited to generating research questions, designing experiments to answer questions, making observations, taking measurements, developing theories, and studying others' research (Chinn & Malhotra, 2001; Osborne, Collins, Ratcliffe, Millar, & Duschl, 2003). While these actions may be encouraged in traditional classrooms, they are typically enacted in experimentation that is simple and fixed which is different from the experimentation of scientists. In contrast, *authentic scientific practice* that scientists engage in involves doing science (1) in the context of real-world problems, (2) where the full range of variables can be tested and the full range of outcomes may be unknown, and (3) where procedures for answering questions are chosen at least partially by participants (Chinn & Malhotra, 2001; Gleason & Schauble, 1999).

Engaging learners in such scientific practice and facilitating the development of a learning community involves many challenges. The difficulties learners face engaging in scientific practice must be addressed, learners' exploration must be in the context of their interests, and collaboration must be promoted for the development of the learning community. Technology can help learners and facilitators overcome some of the challenges by supporting learners' scientific practice and the collaboration of the learning community. Technology can help learners overcome difficulties engaging in scientific practice by scaffolding the processes of inquiry. Specifically, learners need help articulating the important parts of their experiences, making quantitative observations, and making plans for experiments (Clegg & Kolodner, 2007). Quintana et. al., (2004) point out several software systems that have been successful at this type of scaffolding. However, in designing this help for out-of-school environments, the challenge becomes presenting it as an opportunity in the context of helping learners achieve their goals.

Technology support can also promote the development of a learning community in several ways. First, it can help learners build a shared history in their community by allowing them to contribute to the community's repository of shared history in different ways (Scardamalia & Bereiter, 1996). Second, technology support can help learners structure their work, share their ideas, and get feedback from others. This is especially important in a context where learners choose projects to work on and may choose projects that they need help with. Our experiences with promoting learning with this type of software show that we need to give learners both free-form and scientifically-structured means of sharing their ideas and asking for help from the community (Clegg & Kolodner, 2007). Software can also serve to help learners reify their expertise in the community, providing them a platform for pointing out and highlighting their accomplishments and contributions.

While the literature points out ways technology can promote science learning and the development of a learning community, we still need to understand what roles technology can play in a context where learners' motivation determines their participation. We therefore ask in the context of out-of-school environments:

(1) How can technology promote collaborative scientific participation?

(2) What roles can technology play in supporting collaborative scientific participation?

Design of Environment

We have studied these questions in the context of Kitchen Science Investigators (KSI), an out-of-school learning environment we designed for middle-school kids to learn science in the context of cooking. In order to create an environment where participants could learn to use scientific practices and pursue ideas that were personally interesting to them, we designed two activity sequences. In early sessions KSI participants engage in a semi-structured activity sequence to familiarize and scaffold them in engaging in the scientific practices of asking questions, designing experiments, making observations, measuring, sharing results, and drawing conclusions. Once they have become somewhat proficient at carrying out scientific reasoning, participants engage in more flexible exploratory activity sequences where they can use the science they learned during the semi-structured activities to iteratively perfect recipes of their choosing.

The goal of the semi-structured activity sequence is to engage participants in conducting cooking and science experiments that are focused on understanding what makes foods rise or thicken. Together, learners and facilitators design experiments that highlight the effects of varying amounts or types of ingredients in a recipe. The cooking experiments highlight the effects of ingredients (e.g., increased height, volume, density or thickness) in the context of the recipe they are preparing. For example, participants make pudding with different types of starch thickeners and the facilitator helps them to make observations and compare the different textures and thicknesses of the puddings. These experiments are designed as a whole group. In small groups, learners carry out each variation and then share their results, taste their dishes, and draw conclusions as a whole group.

After several semi-structured sessions, learners are given the opportunity to use what they have learned about particular ingredients to prepare recipes of their choice and make them come out with their preferred taste and texture. During these sessions, called choice days, learners ask new questions and practice using results and conclusions drawn from the semi-structured experiments to design experiments that answer their questions. At this point, they are not yet fully proficient at asking questions, using results, drawing conclusions, or designing experiments, but they do know that these are activities they need to engage in to design their own recipes.

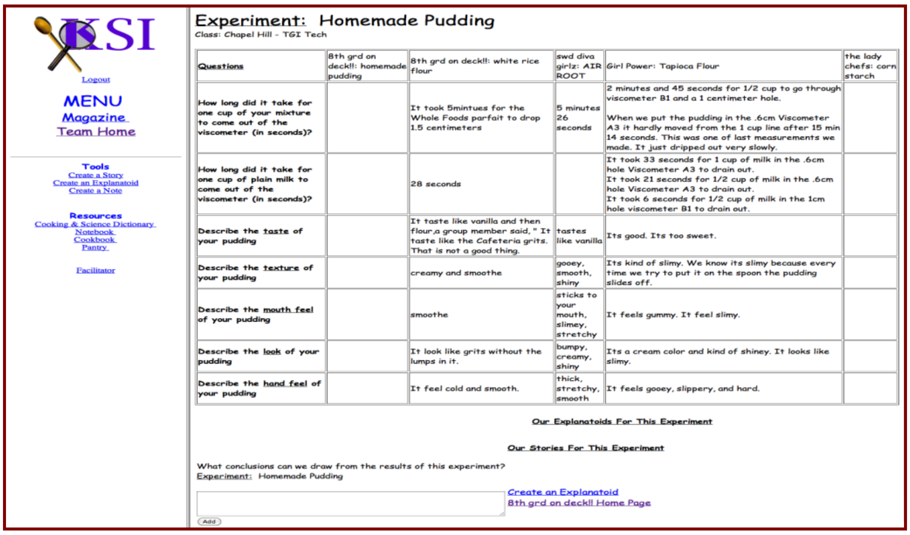
We designed the software in KSI to prompt several aspects of scientific reasoning. We wanted learners working in small groups to fluently use the software during the choice days to record questions, plan experiments, enter data, and draw conclusions. We therefore needed to design it in such a way that it not only supports those activities but it supports them similarly across the contexts of semi-structured and choice days.

During semi-structured days, the software supports: experiment design as learners engage in a large group to design experiments, data collection as they work in small groups, compilation of data collected across groups, and interpretation of that data and question answering by the large group after experimentation is complete. It also supports each of those functions during choice days. However, choice days are more flexible than semi-structured days. The software therefore also needs to support learners in sharing their diverse experiences scientifically.

We designed the software such that it provides the same support across semi-structured and choice day sequences. However, the software offers affordances particularly suited for each type of activity. The software provides support for designing a controlled experiment for each recipe in the system. It prompts learners to formulate questions they want to answer with the experiment, to select one ingredient to vary, to identify dependent variables to measure, and to identify variables they will control to answer their question. The software then guides learners in carrying out the procedures for each variation of the recipe, encouraging them with prompts at each step to make observations. During semi-structured experiments each group carries out one variation of the recipe. During choice days one group may carry out all of the variations of an experiment, or they may only try one variation of a recipe. Once learners have completed their variation(s), the software prompts them with questions about their results that they can enter into the software. These questions are pre-determined by facilitators, and are the questions learners will be able to answer with the particular experiment they designed.

To support learners' scientific understanding of underlying mechanisms (e.g., starch molecule structure and function) in their dishes, we used paper-based scaffolds to provide visualizations and information. We also used paper-based goals charts to help learners describe specific characteristics they wanted their choice day recipes to have. We wanted learners to think of their new, complex dishes in the same terms they described previous experiment results so that they could draw upon their previous experiments more naturally. We used paper-based scaffolding for these aspects of learners' scientific engagement because we created these scaffolds week-to-week specifically based on learners interests and understanding.

Learners could also create and edit stories of their experience and explanatoids (short "Did You Know" facts they discovered or were introduced to while creating their recipes). Stories and explanatoids are particularly useful for enabling learners to share their diverse experiences on choice days. They can then navigate to the experiment results page (Figure 1) where they can view results from all of the variations run in the experiment.



Questions	8th grd on deck!!: homemade pudding	8th grd on deck!!: white rice flour	swd diva girlz: AIR ROOT	Girl Power: Tapioca Flour	the lady chefs: corn starch
How long did it take for one cup of your mixture to come out of the viscometer (in seconds)?		It took 5mintues for the Whole Foods parfait to drop 1.5 centimeters	5 minutes 26 seconds	2 minutes and 45 seconds for 1/2 cup to go through viscometer B1 and a 1 centimeter hole. When we put the pudding in the .6cm Viscometer A3 it hardly moved from the 1 cup line after 15 min 14 seconds. This was one of last measurements we made. It just dripped out very slowly.	
How long did it take for one cup of plain milk to come out of the viscometer (in seconds)?		28 seconds		It took 33 seconds for 1 cup of milk in the .6cm hole Viscometer A3 to drain out. It took 21 seconds for 1/2 cup of milk in the .6cm hole Viscometer A3 to drain out. It took 6 seconds for 1/2 cup of milk in the 1cm hole viscometer B1 to drain out.	
Describe the taste of your pudding		It taste like vanilla and then flour,a group member said, " It taste like the Cafeteria grits. That is not a good thing.	tastes like vanilla	Its good. Its too sweet.	
Describe the texture of your pudding		creamy and smoothe	gooey, smooth, shiny	Its kind of slimy. We know its slimy because every time we try to put it on the spoon the pudding slides off.	
Describe the mouth feel of your pudding		smoothe	sticks to your mouth, slimey, stretchy	It feels gummy. It feel slimy.	
Describe the look of your pudding		It look like grits without the lumps in it.	bumpy, creamy, shiny	Its a cream color and kind of shiney. It looks like slimy.	
Describe the hand feel of your pudding		It feel cold and smooth.	thick, stretchy, smooth	It feels gooey, slippery, and hard.	

Figure 1. Experiment Results Page in the KSI Software.

This page displays a chart that allows learners to compare results across variations, with respect to the questions they asked. Learners' explanatoids and stories from the experiment are listed and can be viewed from this page. As they look across results, they can enter in conclusions they have drawn from their results.

Methods

Recall that the questions we seek to answer in this paper are: (1) How can technology promote collaborative scientific participation? and (2) What roles can technology play in supporting collaborative scientific participation?

The cases presented in this paper were taken from data collected in an enactment of the KSI program, run throughout the 2007-2008 school year. We offered the program afterschool one day a week as part of a larger afterschool initiative by a local YWCA to engage teen girls in science and technology related activities. Participants in the study were from the same suburban middle school where the population was 99% African American. Thus, all participants were African American girls. Participation varied over the 9-month period, but we had 15-20 consistent participants (7-9 6th graders, 7-10 8th graders, one 7th grader). A team of 3 facilitators, the authors of this paper, led the KSI sessions. In every session, video recordings of each group were collected and transcribed. In addition, after each session, facilitators recorded post-observation field notes that captured the significant learning events that occurred during the session.

Data used in this analysis was collected as part of a larger study focusing on which aspects of the learning environment influenced participants' development of scientific reasoning identities. This investigation involved the selection of four focal learners in KSI: Amber, Malaysia, Candyce, and Sharonda. All of the learners were 6th graders, with the exception of Amber, who was an 8th grader. These learners were purposefully selected to represent a range of participation styles and interests.

The first author did a sequence of three in-depth semi-structured interviews with the focal learners. The interviews were structured similar to Seidman's (1991) phenomenological approach. They, however, were spaced out over the second half of the program – to capture learners' change in scientific participation and identity throughout their participation in KSI. Conducting three interviews with each participant allowed us to attain reliability (Seidman, 1991) and to monitor changes in learners' goals over time and the meaning they were making of their experiences in KSI. In order to triangulate the data gathered from the interviews with the learners, we used video observation from their participation in the program, interviews with their parents, and interviews with their science teachers. We conducted initial and ending interviews with focal learners' parents and science teachers to measure the change they saw in the learner. Interviews with parents were focused on learners' scientific participation at home (especially when cooking). Interviews with science teachers were focused on learners' interests, skills, and participation in science class.

Our analysis began with the identification of scientifically meaningful experiences learners had in KSI. Facilitator field notes and learner interviews were coded (and triangulated with parent and teacher interviews) to pinpoint specific scientific experiences in KSI that were meaningful for learners as evidenced by their reporting of them or their reports of extending particular KSI experiences in meaningful ways. Learners' scientific practice was identified by coding data using Chinn & Malhotra's (2001) framework for scientific reasoning. Their framework particularly illustrates what scientific participation looks like as it progresses from simple science typically done in the classroom to authentic scientific participation as practiced by professional scientists. The framework breaks down many of the inquiry processes involved in authentic science investigation (e.g., selecting and controlling variables, planning measures, planning procedures).

Next, we analyzed the set of scientifically meaningful experiences that involved technology use to understand the role the technology played in supporting learners' experiences. Within those experiences there were three themes that emerged: the technology-supported scientifically-meaningful experiences by (1) collating and indexing experiment results, (2) enabling learners and facilitators to make free-form contributions, and (3) serving as a resource for further information. These themes were observed multiple times (typically across multiple groups) and had later impact on learners' experiences (e.g., their reporting of the meaning of the resulting experience in interviews, their later expansion of the activity). We then selected illustrative examples of each of these aspects of technology support. We analyzed each example to understand the context of the use of the technology and the role of the technology in supporting the experience. Validity was established in the discussion of the selected episodes with the second and third authors and by their subsequent review of each episode.

Findings

Eighteen scientifically meaningful experiences were identified across the four focal learners (5 experiences for Amber and Sharonda, 4 experiences for Candyce and Malaysia). Next, we present three cases – one for each aspect of the technology that supported learners' scientifically meaningful experiences. For each case we point out the context of the technology use and the role of the technology in supporting the experience.

Collating and Indexing Experiment Results

The KSI software prompts learners to make observations of their dish once they finish preparing their variation. It then collates the observations of each variation into a chart. Facilitators were able to refer learners to these charts on later days as they made decisions about how to prepare more complex recipes to meet their specific cooking goals. This theme was observed in multiple scientifically meaningful experiences for Candyce.

Context

Day 16 was a choice day where the goal was to create a complex dish using what learners knew about different starch thickeners to make the texture of the dish come out the way they wanted it to. Malaysia and Candyce, two sixth graders, chose to make fruit tarts with Janet (a KSI facilitator). Before they began cooking, they wrote out their goals for their fruit tart on a paper-based goals chart. They wanted their fruit tart to be soft, creamy, moist, and smooth with a sweet and "fruitiliscious" taste. Janet then asked them what thickener they wanted to use in their tart filling. She prompted them to refer back to Day 12's pudding experiment results chart in the software. The chart (Figure 1) presented the results of a pudding experiment they had done as part of the semi-structured activity

They then used the pudding results chart to select the thickeners that produced the pudding that best matched their goals for their fruit tart filling. They decided to use arrowroot because they wanted the taste to be sweet and creamy and white rice flour because they wanted a creamy texture. They were very pleased with their fruit tart's texture and taste, and they were especially excited to take mini tarts home to their families.

In the next session (Day 17), when they re-prepared the fruit tarts using the same thickeners and recipe, their custard became "rubbery." Candyce used her memory of the results from the chart to reason about their unexpected results. She described her reasoning in a later interview, "Um, because when we were reading about the different puddings we saw that arrowroot kind of made it creamy and thick but not too thick. So, I think we added too much of that because of the description." She also later used her memory of this chart as she worked in different groups on Days 18 and 20 to make decisions about what thickener to use in new recipes she made. She based her decisions on her goals for the new recipe and her memory of the results for each thickener variation of pudding.

Role of the Technology

The experiment results chart provided an index of learners' previous experiences. Most often the facilitators pointed participants to those charts at times when they needed to make decisions, but sometimes participants remembered what they had seen in the charts when they had looked at them previously, and sometimes they accessed the charts themselves. Whichever was their way of accessing the data in the charts, use of the data led them to make their decisions based on evidence. Although facilitators pointed other learners and groups back to previous experiment results pages, Candyce's case was particularly meaningful because of the extent to which her group used the results to inform their decision and because of the extent to which she referred back to the results later.

Making Free-form Contributions – Stories and Explanatoids

Facilitators often had to prompt learners to write stories and explanatoids. They also needed to ask learners specific questions about their experiences and type learners' dictation in order to create stories. However, three groups created 7 explanatoids, or simple explanations, and 10 stories were created by 8 groups. Stories and explanatoids were created over 11 sessions of the program and in 4 of the experiences identified as scientifically meaningful. The following example was representative of the way stories were written and used in the learning environment.

Context

On Day 20, Malaysia worked with a larger group of all 8th graders (Amber, Patience, Soleil, and Angela) making fettuccine alfredo (with pasta made from scratch). This was the last day of the program, and after their cooking activity, learners' parents were invited to eat the food learners prepared and to listen to learners' presentations of what they did and learned in KSI.

The group made several observations as they were preparing their pasta. For example, they noticed their pasta became much larger when boiled and that there was white residue in their boiling water after the pasta was removed. The facilitator working with them, Christina, took the opportunity to discuss the science behind some of the things they were observing. She explained that pasta is made of starches (i.e., flour) and how the starches work in making pasta. She used paper-based explanation cards to provide visualizations of different types of starches and how they absorb water. She then connected the changes they were observing in their pasta to the explanations. In preparation for their presentation, Christina helped them create an explanatoid of what they learned as they made their pasta by asking questions about what they had discussed earlier and typing learners' responses (Figure 2).

The Science Behind Pasta

Q. What is pasta made of?
Pasta is made of flour, water, oil, and salt

Q. What is flour made of?
Starch granules and gluten

Q. What are starch granules made of?
Amylose and Amylopectin

Q. How does Amylose work?
Is a starch molecule in starch that doesn't absorb water.

Q. How does Amylopectin work?
Is a starch molecule in starch that doesn't absorb water. Amylopectin is shaped like a branch and the branch structure traps the water.

Q. What is the white sticky stuff in the the pasta pot after you boil pasta?
The white residue is starch. Specifically, it is amylose that doesn't absorb water.

Q. How does the starch get there?
The amylose starches in the pasta came out because they don't absorb water.

Q. What happens when you boil the pasta in water? The pasta expands in the water.

Q. Why does the pasta expand?
The Amylopectin starch molecules they absorb water and swell.

Q. What happen when you over boil pasta?
It will stick and clump up and get gooey.

Figure 2. Part of an explanatoid created on Day 20 about what the group learned as they made pasta. Although this was created in the software as one explanatoid, it actually represents several explanatoids.

Several of the group members left before the presentations began. However, Malaysia, Patience, and Amber stayed to present their experience and dish to their parents. The group began their presentation of the fettuccine alfredo by reading the explanatoid they created to talk about the science behind pasta. Patience asked the questions in their explanatoid while Malaysia and Amber took turns answering them. Their explanatoid (Figure 2) addressed amylose and amylopectin, their differences in absorption of water, and their impacts on pasta.

While the group mainly read their explanatoid, their impromptu corrections and answers to others' questions showed their understanding and familiarity with the scientific concepts they discussed. For example, Malaysia corrected Amber's pronunciation of a science word in the explanatoid – residue. She also knew that gluten is found in flour when Christina asked. During their presentation of their fettuccine alfredo, Tammy (the first author and KSI facilitator) asked, "Well do you think you could make whole grain [pasta]?" Amber affirmed that they could, "it would just be a different flour." Mentioning a different type of flour prompted Malaysia to ask, "Is whole grain that little wheat stuff?" When Amber affirmed Malaysia's question, Malaysia made a connection to branch-shaped amylopectin stating, "It looks like amylopectin, or amylose."

Role of the Technology

The technology combined the voices of the facilitators and learners. The facilitator helped learners share relevant aspects of their experiences in the explanatoid. She also prompted learners to use scientific vocabulary (e.g., amylose, amylopectin). But because she documented the learners' responses to her prompts, their explanatoid (and stories in other cases) were reflective of learners' and facilitators' perspectives of the experience. The learners were then able to use the software to present scientifically relevant aspects of their experiences to the group.

General-Purpose Technology as Resource for Further Information

Another category of technology use emerged that we did not anticipate. Learners and facilitators also used general-purpose technology (e.g., digital cameras, the internet) for their collaborative scientific participation. This was observed in five of the experiences identified as scientifically meaningful. The following example details one particularly compelling case of general technology use to support collaborative scientific inquiry.

Context

On Day 11, Sharonda, Esha, Treeva, Rachel, and Christina (the facilitator) worked together making biscuits from scratch and gravy using store-bought gravy packets. They simply needed to add liquid to the gravy package contents and stir over heat. All participants were sixth graders, except for Rachel, an eighth grader. The groups were transitioning from exploring leaveners (used in biscuits) to thickeners (used in gravy).

Sharonda began preparing the gravy with Christina as her group finished the biscuits. Christina and Sharonda made the gravy and explored the question posed earlier in the whole group discussion: What makes gravy thick? As they stirred the gravy packet contents and water on the stove, Christina encouraged Sharonda to look at the ingredients list on the gravy packet and guess which ingredients thickened the gravy. Sharonda initially thought it was the baking soda in the gravy that made it thick. However, she continued to read the ingredients list on the packet and ask Christina about them. When Christina did not know what one ingredient was, she suggested an online search. When Christina was called away, she told Sharonda where to look online for the definition and left. Once Sharonda found the first word, she called out, "Ms. Christina, I found a definition!"

When Christina returned, they read the definition and Christina asked, "So do you think this is gonna thicken it?" Sharonda replied, "no" and they continued to search for the ingredients on the online dictionary. As Sharonda read the definitions, she came across some complex words and definitions (e.g., folic acid). Christina helped her to interpret the meaning of each and keep track of which ingredients they had looked up that could possibly serve to thicken the gravy. When Sharonda's science teacher (who was present) asked them about what they were doing, Christina explained, "We're trying to figure out what made our gravy thick, in the packet." Sharonda added, "And we're having fun." Later, Sharonda discussed how the experience was useful to her:

> Um, it really…it told us about the words. It helped us um ... know what the words mean so when we go home --if our parents ask us about a word we just looked up, we would know what it means.

Role of the Technology

In this case, the Internet supported a one-on-one search between the learner and facilitator. The facilitator needed to be there to help the learner with the difficult language and concepts and to help her think through the function of each of the ingredients. The technology provided information Christina, the facilitator, did not know, and it enabled her to model the process of searching for answers to questions to the learner. The technology also helped Christina multitask between the investigation and her responsibilities with other groups. When she had to step away from the investigation, Sharonda carried out more searches while she was gone, then discussed each term and definition she found with Christina when she returned. The technology also helped position Sharonda as an expert.

Discussion

We found that the software never became as central a part of the environment as we expected. Throughout the semi-structured sequence and exploratory activities, the cooking activities remained the focal point of learners' activity. However, even though the software remained in the background of the learning environment and activities, it still played an essential role in promoting collaborative scientific participation. We now return to our research questions:

How does technology promote collaborative scientific participation? We found that because the technology acted as a repository of learners' experiences, it scaffolded having and sharing scientific experiences while cooking. We designed the scaffolding to be used directly by learners. We're guessing that if KSI had progressed, learners would have eventually taken initiative to use the resources themselves. In the meantime, however, having that function there allowed the facilitator an easy time helping learners draw on their scientific experiences to make decisions for new recipes, create scientific artifacts of their cooking experiences that they could share with others, and find scientific answers to their own questions about the recipes they made.

The technology also facilitated collaboration between the facilitators and learners. It did so by supporting the roles of both learners and facilitators. The structured software support helped facilitators help learners match their cooking goals with previous experiment results. It also helped facilitators by providing an easily accessible tool to point learners to in helping them cook scientifically. Creating explanatoids (and stories) helped facilitators help learners highlight scientific aspects of their experiences. The presence of the function in the software, even though it was the facilitator who did the typing, allowed learners to present their experiences to others. Finally, general-purpose technology use supported learners and facilitators in answering their own questions.

What roles can technology play in supporting collaborative science participation? The technology in KSI played two essential roles in the learning environment. First, it helped connect science to the context of cooking. The experiment results page compiled each group's cooking variation into a chart that allowed learners to compare across variations. This chart then helped learners to compare their later cooking goals to their experimental results. Creating stories and explanatoids helped learners to share their cooking experiences scientifically with others – using scientific terminology and presenting relevant aspects of their experiences. Use of the Internet enabled learners and facilitators to explore scientific aspects of the dishes they were creating.

The second role the technology played in the learning environment was documenter, or recorder of experiences. Although it never played a central role in the learning environment, the software kept track of learners experiment results as well as the scientific discussions learners had during their cooking experiences. Its presence allowed facilitators to easily point back to those artifacts for use of the data to make decisions in new recipes and for sharing scientific experiences and understanding with others.

Although the software was not always at the forefront in the learning environment, its presence enabled it to play essential roles in facilitating learners' scientific participation as well as their collaboration with facilitators. In other CSCL work, the software plays a central role in the learning environments they are situated within. Learners use the technology to communicate and it drives the activity (e.g., Hickey et. al., 2003; Reiser et al., 2001) or the software functions drive conceptual engagement as learners use it together (Dasgupta & Kolodner, 2009; Scardamalia & Bereiter, 1996). Our work makes a contribution to the CSCL community in showing that technology does not have to play a central role in the learning environment to be important and effective for scaffolding learners and supporting collaboration. Specifically, we have shown that when software is used to prompt science connections and to record experiences and understanding, it can promote collaborative engagement in science talk and scientific reasoning by learners and promote productive interactions between facilitators and learners.

On the other hand, the sequencing of learning activities and facilitation support *were* central aspects of the learning environment. Analysis of the role of the facilitators and activities in KSI is beyond the scope of this paper. Yet, it is important to acknowledge in considering the role of technology in environments supporting scientific collaboration in the context of learners' interests and goals. Learning activities sequenced from more to less structured all in the context of learners' goals were important for sustaining motivation and promoting scientific participation. Facilitators then played critical roles in helping learners focus on scientific aspects of their experiences, understanding relevant scientific phenomena, and engaging in scientific practice. At times, their interactions with learners resembled that of teachers in traditional classrooms. However, unlike teachers, facilitators were not bound to a set curriculum, they were bound to learners' interests, curiosities, and questions. Technology therefore needs to support such collaboration between facilitators and learners by providing the structure needed to scaffold science learning and the freedom needed to document, share, and learn from diverse, dynamic experiences.

Our work is limited in that it presents data from a small sample of learners in one learning environment. The technology we presented does not provide any new types of computer support, and some of the support was quite simple (e.g., providing space to write stories). We did not have to implement any fancy functions to effectively support participants' and facilitators' collaborative engagement in scientific practices. The key was integrating use of the software into the complex physical environment (Roschelle, 2003). The facilitators had to play a critical role in getting the kids to use the software. However, once used, it provided an organized repository of the community's experiences and put information at learners' and facilitators' fingertips (via the Internet) that simple paper and posters could not have easily provided.

References

Chinn, C. A., & Malhotra, B. A. (2001). Epistemologically authentic scientific reasoning. In K. Crowley, C. D. Schuunn & T. Okada (Eds.), *Designing for science: Implications from everyday classroom, and professional settings* (pp. 351-392). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Clegg, T., & Kolodner, J. (2007). Bricoleurs and planners engaging in scientific reasoning: A tale of two groups in one learning community. *Research and Practice in Technology Enhanced Learning, 2*(3), 239-266.

Crowley, K., & Jacobs, M. (2002). Building islands of expertise in everyday family activity. In G. Leinhardt, K. Crowley & K. Knutson (Eds.), *Learning conversations in museums*. Mahway, NJ: Lawrence Erlbaum Associates.

Dasgupta, C., & Kolodner, J. L. (2009). *Designing case libraries to encourage creative design*. Paper presented at the Creativity and Cognition, Berkeley, CA.

Gardner, C. M., Clegg, T. L., Williams, O. L., & Kolodner, J. L. (2006). *Messy learning environments: Busy hands and less engaged minds.* Paper presented at the Seventh International Conference of the Learning Sciences Bloomington, IN.

Gleason, M. E., & Schauble, L. (1999). Parents' assistance of their children's scientific reasoning. *Cognition and Instruction, 17*(4), 343 - 378.

Osborne, J., Collins, S., Ratcliffe, M., Millar, R., & Duschl, R. (2003). What "ideas-about-science" should be taught in school science? A delphi study of the expert community. *Journal of Research in Science Teaching, 40*(7), 692-720. doi: 10.1002/tea.10105

Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., . . . Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. *Journal of the Learning Sciences, 13*(3), 337-386.

Roschelle, J. (2003). Keynote paper: Unlocking the learning value of wireless mobile devices. *Journal of Computer Assisted Learning, 19*(3), 260.

Scardamalia, M., & Bereiter, C. (1996). Computer support for knowledge-building communities. In T. Koschmann (Ed.), *Cscl: Theory and practice of an emerging paradigm* (pp. 249-268). Mahwah, NJ: Lawrence Erlbaum Associates.

Seidman, I. E. (1991). *Interviewing as qualitative research: A guide for researchers in education and the social sciences*. New York, NY: Teachers College Press.

How a Virtual Math Team Structured its Problem Solving

Gerry Stahl, Drexel University, Philadelphia, PA, USA, Gerry@GerryStahl.net

Abstract: To develop a theory of small-group interaction in CSCL settings, we need an approach to analyzing the structure of computer-mediated discourse. Conversation Analysis examines informal face-to-face talk in terms of a fine structure of adjacency pairs, but needs to be adapted to online textual interaction and extended to analyze longer sequences built on adjacency pairs. This paper presents a case study of students solving a math problem in an online chat environment. It shows that their problem-solving discourse consists of a sequence of exchanges, each built on a base adjacency pair and each contributing a move in their collaborative problem-solving process.

Structuring Group Cognition at Multiple Levels

A year ago in my opening keynote talk (Stahl, 2009a) at the International Conference of Computers in Education (ICCE 2009) in Hong Kong, I claimed that the discourse of group cognition (Stahl, 2006) has a hierarchical structure, typically including the following levels, as illustrated with a particular case study from the Virtual Math Teams (VMT) Project (Stahl, 2009c):

a. **Group event**: E.g., Team B's participation in the VMT Spring Fest 2006.
b. **Temporal session**: Session 4 of Team B on the afternoon of May 18, 2006.
c. **Conversational topic**: Determining the number of sticks in a diamond pattern (lines 1734 to 1833 of the chat log of Session 4).
d. **Discourse move**: A stage in the sequence of moves to accomplish discussing the conversational topic (e.g., lines 1767-1770—see Logs 1-10 below).
e. **Adjacency pair**: The base interaction involving two or three utterances, which drives a discourse move (lines 1767 and 1769).
f. **Textual utterance**: A text chat posting by an individual participant, which may contribute to an adjacency pair (line 1767).
g. **Indexical reference**: An element of a textual utterance that points to a relevant resource. In VMT, actions and objects in the shared whiteboard are often referenced in the chat. Mathematical content and other resources from the joint problem space and from shared past experience are also brought into the discourse by explicit or implicit reference in a chat posting.

The multi-layered structure corresponds to the multiplicity of constraints imposed on small-group discourse—from the character of the life-world and of culture (which mediate macro-structure) to the semantic, syntactic and pragmatic rules of language (which govern the fine structure of utterances). A theory of group cognition must concern itself primarily with the analysis of mid-level phenomena—such as how small groups accomplish collaborative problem solving and other conversational topics.

The study of mid-level group-cognition phenomena is a realm of analysis that is currently underdeveloped in the research literature. For instance, many CSCL studies focus on coding individual (micro-level) utterances or assessing learning outcomes (macro-level), without analyzing the group processes (mid-level). Similarly, Conversation Analysis (CA) centers on micro-level adjacency pairs while socio-cultural Discourse Analysis is concerned with macro-level identity and power, without characterizing the interaction patterns that build such macro phenomena out of microelements. Understanding these mid-level phenomena is crucial to analyzing collaborative learning, for it is this level that largely mediates between the interpretations of individuals and the socio-cultural factors of communities.

The analysis in this paper illustrates the applicability of the notion of a 'long sequence' as vaguely suggested by both Sacks (1962/1995, II p. 354) and Schegloff (2007, pp. 12, 213). A longer sequence consists of a coherent series of shorter sequences built on adjacency pairs. This multi-layered sequential structure will be adapted in this paper from the informal face-to-face talk-in-interaction of CA to the essentially different, but analogous, context of groupware-supported communication and group cognition, such as the text chat of VMT. I will show how a small group of students collaborating online constructed a coherent long sequence, through which they solved the problem that they had posed for themselves. Methodologically, it is important to note that the definition of the long sequence—like that of the other levels of structure listed above—is oriented to by the discourse of the students and is not simply a construct of the researcher.

An Analytic Method

Recently, I have been trying to apply the CA perspective and techniques in a systematic way to the analysis of VMT chat logs. Schegloff's (2007) book on *Sequence Organization in Interaction* represents the culmination of decades of CA analysis. As indicated by its subtitle, it provides a useful primer in CA. My goal is to transform CA to apply to online chat and to extend it to analyze the larger scale interactions of group cognition.

Schegloff's presentation makes clear the central role of the adjacency pair as the primary unit of sequence construction according to CA. An adjacency pair is composed of two conversational speaking turns by two different people, with an interactional order, such as a question followed by an answer to the question. The simple two-turn pair can be extended with secondary adjacency pairs that precede, are inserted between or follow up on the base pair, recursively. This yields "extensive stretches of talk which nonetheless must be understood as built on the armature of a single adjacency pair, and therefore needing to be understood as extensions of it" (p. 12).

These "extensive stretches of talk" are still focused on a single interaction of meaning making, and not a larger cognitive achievement like problem solving. However, both Sacks and Schegloff provide vague suggestions about the analysis of longer sequences. These suggestions have not been extensively developed within CA. This paper is an attempt to explore them in an online text-chat context.

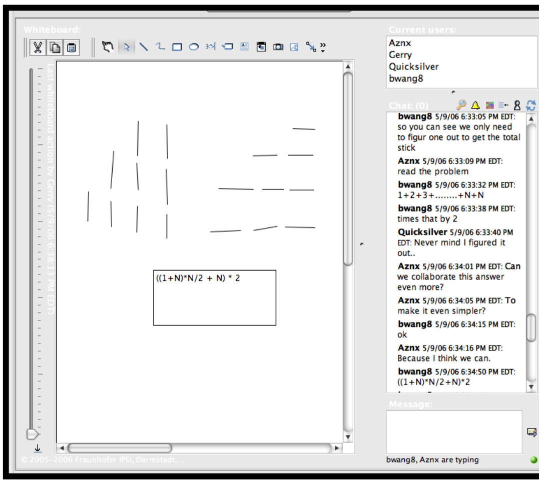


Figure 1. VMT interface with stair-step pattern of horizontal and vertical sticks.

As I have frequently argued (e.g., Stahl, 2006; 2009c; Stahl, Koschmann & Suthers, 2006), I believe that adapting CA to computer-mediated communication offers the best prospects for analysis of interaction in groupware—i.e., for a theory of small groups appropriate to CSCL. I designed and directed the Virtual Math Teams (VMT) Project from 2003 to the present in order to produce a corpus of data that could be analyzed in as much detail as needed to determine the structure of group cognition, that is, of collaborative knowledge building through interaction at the group unit of analysis.

In looking at the VMT data corpus, the VMT research team has clearly seen the differences between online text chat and verbal conversation. The system of turn taking so important in CA (Sacks, Schegloff & Jefferson, 1974) does not apply in chat. Instead, chat participants engage in 'reading's work' (Zemel & Çakir, 2009), in which "readers connect objects through reading's work to create a 'thread of meaning' from the various postings available for inspection" (p. 274f). The first and second parts of an adjacency pair may no longer be literally temporally adjacent to each other, but they still occur as mutually relevant, anticipatory and responsive. The task of reading's work—for both participants and analysts—is to reconstruct the threading of the adjacency pair response structure (Stahl, 2009b).

We have tried to explore the larger sequential structure of problem-solving chat by using the CA notion of openings and closings (Schegloff & Sacks, 1973). VMT researchers looked at several math chats from 2004, which used a simple chat tool from AOL. We coded and statistically analyzed the fine-structure threading of adjacency pairs (Çakir, Xhafa & Zhou, 2009). In addition, we defined long sequences based on when opening and closing adjacency pairs achieved changes in topic (Zemel, Xhafa & Çakir, 2009). These long sequences were graphed to show their roles in constituting the chat sessions, but their internal sequential structures were not investigated.

My colleagues and I have subsequently conducted numerous case studies from the VMT corpus. We have been particularly drawn to the records of Team B and Team C in the VMT Spring Fest 2006. These were particularly rich sessions of online mathematical knowledge building because these teams of students met for over four hours together and engaged in detailed explorations of interesting mathematical phenomena. However, partially because of the richness of the interactions, it was often hard for analysts to determine a clear structure to the student interactions. Despite access to everything that the students knew about each other and about the group interaction, it proved hard to unambiguously specify the group-cognition processes at work (Medina, Suthers & Vatrapu, 2009; Stahl, 2009b; Stahl, Zemel & Koschmann, 2009).

Therefore, in the following case study, I have selected a segment of Team B's final session, in which the structure of the interaction seems to be clearer. The interaction is simpler than in earlier segments partially because two of the four people in the chat room leave. Thus, the response structure is more direct and less interrupted. In addition, the students have already been together for over four hours, so they know how to interact in the software environment and with each other. Furthermore, they set themselves a straightforward and well-understood mathematical task. The analysis of this relatively simple segment of VMT interaction can then provide a model for subsequently looking at other data and seeing if it may follow similar patterns.

The Case Study

Three anonymous students (Aznx, Bwang, Quicksilver) from US high schools met online as Team B of the VMT Spring Fest 2006 contest to compete to be "the most collaborative virtual math team." They met for four hour-long sessions during a two-week period in May 2006. A facilitator (Gerry) was present in the chat room to help with technical issues, but not to instruct in mathematics.

In their first session, they solved a given problem, finding a mathematical formula for the growth pattern of the number of squares and the number of sticks making up a stair-step arrangement of squares. They determined the number of sticks by drawing just the horizontal sticks together and then just the vertical ones (see Figure 1). They noticed that both the horizontals and the verticals formed the same pattern of **1 + 2 + 3 + … + n + n** sticks at the $\mathbf{n}^{th}$ stage of the growth pattern. They then applied the well-known Gaussian formula for the sum of consecutive integers, added the extra **n**, and multiplied by **2** to account for both the horizontal and vertical sets of sticks.

In the second session, they explored problems that they came up with themselves, related to the stair-step problem, including 3-D pyramids. Here they ran into problems drawing and analyzing 3-D structures. However, they managed to approach the problem from a number of perspectives, including decomposing the structure into horizontal and vertical sticks. In the third session, Team B was attracted to a diamond-shaped variation of the stair-step figure, as explored by Team C in the Spring Fest. They tried to understand how the other team had derived its solution. They counted the number of squares by simplifying the problem through filling in the four corners surrounding the diamond to make a large square; the corners turned out to follow the stair-step pattern from their original problem. In the fourth session, they discovered that the other team's formula for the number of sticks was wrong. In the following, we join them an hour and 17 minutes into the fourth session, when one of the students as well as the facilitator had to leave.

Problem-Solving Moves

In this section of the paper, the interaction is analyzed as a sequence of moves in the problem-solving interaction between Bwang and Aznx, the two remaining students. Each move is seen to include a base adjacency pair (in **bold** face), which provides the central interaction of the move and accomplishes the focal problem-solving activity. The captions of log segments indicate the aim of the move, according to the analysis.

In line 1734 (see Log 1), Bwang states that the team is close to being able to solve the problem of the number of sticks in the $\mathbf{n}^{th}$ stage of the diamond pattern, suggesting that they might stay and finish it up. Note that this is the end of the last of the scheduled four sessions for the contest, despite some arrangements underway to allow the team to continue to meet. Aznx responds in line 1736, indicating—and implicitly endorsing the suggestion—that the team could indeed continue to work on the current topic. This opens the topic for the group. Quicksilver apologetically stresses that he must leave immediately. He just wants to know the location of the new chat room that the facilitator is setting up for the team to continue its math explorations on a future date. The facilitator supplies this information and everyone says goodbye to Quicksilver.

Log 1. Open a Topic.

LINE	TIME	AUTHOR	TEXT OF CHAT POSTING
1734	**08.17.20**	**bwang8**	**i think we are very close to solving the problem here**
1735	08.17.35	Quicksilver	Oh great...I have to leave
1736	**08.17.39**	**Aznx**	**We can solve on that topic.**
1737	08.17.42	Quicksilver	Sorry guys
1738	08.17.45	bwang8	oh
1739	08.17.46	Aznx	It shouldn't take much time.
1740	08.17.47	bwang8	ok
1741	08.17.50	Aznx	k, bye Quicksilver
1742	08.17.52	Quicksilver	Just tell me the name of the room
1743	08.17.52	bwang8	bye
1744	08.18.14	Gerry	The new room is in the lobby under Open Rooms
1745	08.18.44	Gerry	It is under The Grid World. It has your names on it
1746	08.18.49	Quicksilver	[leaves the room]
1747	08.19.00	Aznx	Alright found it.
1748	08.19.04	Aznx	Thanks.

Log 2. Decide to Start.

1749	08.19.12	Aznx	I guess we should leave then.
1750	**08.19.34**	**bwang8**	**well do you want to solve the problem**
1751	08.19.36	bwang8	i mean
1752	08.19.39	bwang8	we are close
1753	**08.19.48**	**Aznx**	**Alright.**
1754	08.19.51	bwang8	i don't want to wait til tomorrow
1755	08.19.53	bwang8	ok

Aznx expresses uncertainty about how to proceed now that Quicksilver has gone and the facilitator has arranged things for the future. Line 1749 (see Log 2) questions whether he and Bwang need to go as well. Bwang then reiterates his suggestion that they could stay and finish solving the problem. He argues that it should not take much longer. Bwang directly asks Aznx if he wants to solve the problem now. Aznx agrees by responding to Bwang's question in the affirmative. This effects a decision by the pair of students to start working on the problem right away. Bwang continues to argue for starting on the problem now—posting line 1754 just 3 seconds after Aznx' agreement, probably just sending what he had already typed before reading Aznx' response. Bwang then notes the response.

Log 3. Pick an Approach.

1756	**08.19.55**	**Aznx**	**How do you want to approach it?**
1757	**08.20.14**	**bwang8**	**1st level have 1*4**
1758	08.20.20	Gerry	You can put something on the wiki to summarize what you found today
1759	08.20.29	bwang8	2st level have (1+3)*4
1760	08.20.32	Aznx	bwang you put it.
1761	08.20.35	Aznx	for the wiki
1762	08.20.37	bwang8	ok
1763	08.20.42	Aznx	we actually did quite a lot today
1764	08.20.53	bwang8	3rd level have (1+3+5)*4
1765	08.21.05	bwang8	4th level have (1+3+5+7)*4
1766	08.21.10	Gerry	This is a nice way to solve it

Log 4. Identify the Pattern.

1767	**08.21.12**	**Aznx**	**So it's a pattern of +2s?**
1768	08.21.15	Aznx	Ah ha!
1769	**08.21.15**	**bwang8**	**yes**
1770	08.21.20	Aznx	There's the pattern!

Log 5. Seek the Equation.

1771	08.21.39	bwang8	now we have to find a equation that describe that pattern
1772	08.21.49	Aznx	Hold on.
1773	08.21.51	Aznx	I know it.
1774	**08.21.57**	**bwang8**	**what is it**
1775	08.21.58	Aznx	But I'm trying to remember it. =P
1776	08.22.04	Aznx	and explain it as well.
1777	08.22.17	Aznx	try and think of it
1778	08.22.53	Gerry	Maybe Quicksilver can come back here tomorrow or next week to finish it with you
1779	08.23.01	Gerry	I have to go now
1780	08.23.05	Gerry	Bye!
1781	08.23.06	bwang8	ok
1782	08.23.07	bwang8	bye
1783	08.23.23	Gerry	[leaves the room]
1784	08.23.29	bwang8	ok
1785	08.23.32	bwang8	so
1786	08.23.37	bwang8	i think it is this
1787	08.23.53	Aznx	ok
1788	08.23.55	Aznx	i found it
1789	**08.24.00**	**Aznx**	**n^2**
1790	08.24.01	bwang8	(2*n)*n/2
1791	08.24.09	Aznx	or (n/2)^2

Once a decision has been made to solve the problem, the question of how to approach the problem is raised in line 1756. Bwang immediately lays out his approach in lines 1757, 1759, 1764 and 1765 of log 3. The approach is the same as they used in the first session: visualize just the vertical or just the horizontal sticks. The two sets follow the same pattern. In fact, the diamond is also symmetric left/right and top/bottom, so the vertical sticks can be divided left/right into two identical sets, which can then be divided top/bottom. This produces four sets of sticks, each having rows of **1, 3, 5, 7, …** sticks, up to **(2n-1)** for the $\mathbf{n}^{th}$ stage of the diamond pattern. Interspersed with this defining of the approach is a parting reminder from the facilitator, before he logs out, to summarize the team's work on the Spring Fest wiki for other teams to view.

Aznx has previously been oriented toward finding patterns of growth in the mathematical objects the group has been exploring. Often, someone will create a graphical representation of the object in such a way that it makes the pattern visible. Aznx will then formulate a textual description of the pattern. Then the group will work on a symbolic representation to capture the pattern in a mathematical formula. (See (Çakır, Zemel & Stahl, 2009) for an analysis of the intertwining of graphical/visual, textual/narrative and symbolic/mathematical modes of interaction within the work of Team C.) In log 4, line 1767, Aznx describes the pattern as involving adding numbers that successively increase by **2**. The number of sticks in a given stage of the diamond shape is a sum of numbers that start at **1** and increase successively by **2**. When going from one stage to the next, one simply adds another number to this sum that is **2** more than the highest previous one. Aznx presented his description as a question and Bwang affirmed it at the same time as Aznx posted line 1768. Aznx then emphasized that they had identified the pattern

In log 5, Bwang indicates that the next step in their work is to "find an equation that describes the pattern." Aznx asks Bwang to let him state the equation, implicitly agreeing that this is the next step by trying to produce the equation. In line 1774, Bwang asks Aznx to state the equation and Aznx expresses difficulty in formulating an adequate and accountable answer. After a half minute of silence with still no formulation from Aznx, the facilitator suggests that Aznx and Bwang might want to wait until a future time when the whole group can work together to finish the problem. The facilitator then says goodbye and leaves the room. After more than a minute since Aznx posted anything, Bwang starts to preface the presentation of his own formulation. Eventually, Aznx joins back in. Simultaneously, Aznx and Bwang post their formulae. For Aznx, it is either $\mathbf{n^2}$ or $\mathbf{(n/2)^2}$. For Bwang, it is **2n(n/2)**. Aznx has not given any indication of how he got his proposed formulae. Bwang's formula suggests the use of Gauss' summation, which the students have used repeatedly in the past. According to this summation of an arithmetic sequence of integers, the result is the sum of the first and last member of the sequence times half the number of members. For a sequence of **n** members, **1 + 3 + 5 + … + (2n-1)**, the sum would be **[1 + (2n-1)]*(n/2)**. Adding the **1** and the **-1**, yields Bwang's

Log 6. Negotiate the Solution.

1792	08.24.14	Aznx	I'm simplifying
1793	08.24.30	Aznx	if u simplify urs
1794	**08.24.35**	**Aznx**	**its n^2**
1795	08.24.59	Aznx	bwang
1796	08.25.01	Aznx	you there?
1797	**08.25.03**	**bwang8**	**so that's wrong**
1798	08.25.07	bwang8	yeah
1799	08.25.08	bwang8	i am here

Log 7. Check Cases.

1800	08.25.11	Aznx	so
1801	08.25.13	Aznx	the formula
1802	**08.25.22**	**Aznx**	**would be 4n^2?**
1803	08.25.28	bwang8	let's check
1804	08.25.55	bwang8	Yes
1805	**08.26.00**	**bwang8**	**it actually is**
1806	08.26.02	Aznx	So we got it!

Log 8. Celebrate the Solution.

1807	08.26.02	bwang8	omg
1808	08.26.04	Aznx	yay!
1809	**08.26.08**	**bwang8**	**i think we got it!!!!!!!!!!!!**
1810	**08.26.12**	**Aznx**	**WE DID IT!!!!!!**
1811	08.26.12	bwang8	and it is so simple
1812	08.26.14	Aznx	YAY!!!!
1813	08.26.16	Aznx	i know
1814	08.26.17	bwang8	lol
1815	08.26.18	Aznx	lol

Log 9. Present a Formal Solution.

1816	**08.26.34**	**Aznx**	**So you're putting it in the wiki, right?**
1817	**08.26.37**	**bwang8**	**yes**
1818	08.26.41	Aznx	Alright then.
1819	08.26.43	bwang8	ok
1820	08.26.53	Aznx	Give an email to Gery, telling him that we got it. =)
1821	08.26.57	bwang8	ok
1822	08.26.59	Aznx	I meant Gerry
1823	08.27.04	bwang8	are you going to do it
1824	08.27.07	bwang8	or am i
1825	08.27.12	Aznx	You do it.
1826	08.27.14	bwang8	ok
1827	08.27.19	Aznx	Tell him that we both dervied n^2
1828	08.27.29	Aznx	And then we saw that pattern
1829	08.27.37	Aznx	and we got the formula

Log 10. Close the Topic.

1830	08.27.44	Aznx	when should we meet again?
1831	08.27.49	Aznx	hat's your email?
1832	**08.27.52**	**Aznx**	**we should keep in touch**
1833	**08.27.57**	**bwang8**	**yeah**

formula, **2n(n/2)**. Note that the **n**th odd integer can be represented by **(2n-1)**. It is likely that Aznx used a similar method, working on his own during his prolonged silence, but got confused about the result when he simplified his expression. As Aznx shows next, Aznx's first answer is equivalent to Bwang's answer, once Aznx simplifies it. His second answer is related to part of Bwang's unsimplified answer.

Aznx simplifies Bwang's formula: $\mathbf{2n(n/2) = n^2}$ in line 1794 of log 6. This is the same as one of Aznx' proposed formulae. When Bwang does not respond to this posting, Aznx wonders if Bwang is still present online. Bwang was apparently already typing "so that is wrong" when be received Aznx' question concerning his presence. This message in effect confirmed that Aznx' second formula, $\mathbf{(n/2)^2}$, is wrong and his first one, which agrees with Bwang's, is correct. The selection of the solution is thereby negotiated.

Going along with this in line 1802 of log 7, Aznx then multiplies their agreed upon formula by **4** because there were 4 sets of horizontal or vertical sticks, each numbering **1 + 3 + …**. Aznx poses his message as a question, soliciting confirmation from Bwang. By offering this next step in the symbolic representation, Aznx demonstrates that he understands where Bwang's formula came from and he understands the larger strategy of approaching the problem that Bwang had proposed. In other words, Aznx demonstrates a level of mathematical competence and of shared understanding that he did not always display in the previous sessions. Before being ready to answer whether $\mathbf{4n^2}$ is actually the correct formula for the number of sticks, Bwang suggests that they first check if the formula works by testing it for a number of values of **n** and counting the sticks in drawings of diamonds at the corresponding **n**th stage. A half-minute later, Bwang concludes that the formula does check out. He therefore answers Aznx' question with confidence, perhaps mixed with a touch of surprise. Aznx concludes that they got the solution for the number of sticks in the diamond pattern—a problem that Team C had posed for itself, but for which they had derived the wrong formula, without, however, realizing it. Team B had been shocked earlier to discover that the formula they had been struggling to understand from Team C had been wrong; that it did not check out for any values of **n**.

Their surprise and excitement is almost uncontrollable. They use every chat technique they know to express their joy in log 8. Their postings intertwine like a frenzied dance.

Once the mathematical exploration is done, it is time to write up a report of ones findings. Professional mathematicians would do this in the form of a proof. Bwang agrees in log 9 to post a narrative of their solution to the Spring Fest wiki.

Finally, in log 10, Aznx and Bwang wrap up the conversational topic by exchanging email addresses and agreeing to meet again online with Quicksilver and pursue further mathematical adventures together.

The Sequence of Pairs

Within each of the preceding log segments we have identified a base adjacency pair by means of which the work of a specific move in the problem-solving effort of the small group is interactively accomplished. In most cases, a question is posed and a response is then given to it. As Schegloff (2007) argues, an adjacency pair is itself a sequence. It embodies a temporal structure, with the first element of the pair projecting the opportunity and expectation of a response in the interactional immediate future. The second element constitutes an uptake of a first element that it implicitly references as in the interactional immediate past (Suthers et al., 2010). In engaging in the exchange of an adjacency pair, the participants in the interaction effectively co-construct an elementary temporal structure within which future and past events can be located and referenced.

In talk-in-interaction, as analyzed by CA, the immediacy of response is intimately related to the turn-taking structure of vocal conversation (Sacks, Schegloff & Jefferson, 1974). As discussed above, the completion of the adjacency pair can be postponed by insertion sequences, such as repairs of misunderstandings or clarification exchanges. The base adjacency pair can also be preceded by introductory exchanges, such as announcements of what is coming, or succeeded by follow-up exchanges or confirmations.

In chat-in-interaction, as seen in the preceding log extracts, adjacency pairs can in addition be delayed by a more complicated response structure, in which multiple participants can be typing simultaneously and postings do not always directly follow the message to which they are responding. Thus, in Log 1, Quicksilver or Gerry can be initiating other topics in the midst of an interaction between Aznx and Bwang. Also, Aznx and Bwang can be typing to each other simultaneously as in Log 6, particularly if there has been an extended period of inactivity. This often makes textual chat harder to follow and to analyze than verbal conversation.

Nevertheless, it is generally possible to identify base adjacency pairs carrying the discourse along. In the previous section, we identified ten such pairs. The discourse moves in the log segments (each including one of these base adjacency pairs) formed a problem-solving sequence:

- Log 1. Open the topic
- Log 2. Decide to start
- Log 3. Pick an approach
- Log 4. Identify the pattern
- Log 5. Seek the equation
- Log 6. Negotiate the solution
- Log 7. Check cases
- Log 8. Celebrate the solution
- Log 9. Present a formal solution
- Log 10. Close the topic

The integrity of each of the ten moves is constructed by the discourse of the participants. Each move contains its single adjacency pair, which drives the interaction. In addition, there may be several utterances of secondary structural importance, which introduce, interrupt or extend the base pair; there may also be some peripheral utterances by other participants.

The Group Perspective

The analysis of this paper is an attempt to make explicit the structure of adjacency pairs and a problem-solving longer sequence that is experienced by the participants and is implicit in the formulation of their contributions to the discourse. This is in contrast to analytic approaches that to some degree impose a set of coding categories based on the analyst's research interests or on an a priori theoretical framework.

Lines 1795 and 1796, for instance, show the power *for the participants* of the adjacency pairings. Here, Aznx has addressed a mathematical proposal to Bwang: "**If you simplify yours** [expression]**, it is n^2.**" After 24 seconds of inaction, Aznx cannot understand why Bwang has not replied, expressing agreement or disagreement with the first part of the proposal, for which Aznx expects a response. Because it is not a preferred move at this point for Aznx to reprimand Bwang for not responding, Aznx inquires if Bwang has disappeared, perhaps due to a technical software problem, which would not be anyone's fault. Two seconds later, we see that Bwang was typing a more involved response that implicitly accepted Aznx' proposal. Bwang then immediately explicitly accepts the proposal in line 1798, allowing Aznx to continue with the start of a new move with line 1802. Here we see Aznx and Bwang clearly orienting to the adjacency-pair structure of their discourse, in terms of their expectations and responses.

Aznx and Bwang co-constructed the longer (ten-move) problem-solving sequence by engaging in the successive exchange of adjacency pairs. Sometimes one of the students would initiate a pair, sometimes the other. As soon as they completed one pair, they would start the next. This longer sequence also has a temporal structure. It is grounded in their *present* situation, trying to find a formula for the number of sticks in the diamond figure. It makes considerable use of resources from their shared (co-experienced) *past* during the previous four hours of online sessions. It is strongly driven forward into the *future* by the practices they have learned for engaging in problem solving, culminating teleologically in the presentation of a solution.

The problem-solving sequence analyzed in this paper—covering 100 lines of chat during 10 minutes—is not selected arbitrarily or imposed in accordance with criteria external to the interaction, but is grounded in the discourse as structured by the participants. The excerpted sequence is defined as a coherent conversational topic *by the discourse* of Aznx and Bwang. They open this topic with their interaction in Log 1 and they close it with the discourse move in Log 10 (Schegloff & Sacks, 1973).

In this paper, I have shown how the group constructs its mid-level problem-solving structure through a longer sequence built on micro-level adjacency pairs and contributing significantly to their macro-level collaborative learning, knowledge building and group cognition. Reviewing the hierarchy of levels introduced previously, it is now clear that each level can be analyzed as *oriented to* by the group discourse and the contributions to it by individual members:

a. **Group event**: The group members log in to the event and comment explicitly on its goals, characteristics and duration as an event in which they are participating.
b. **Temporal session**: Participants start each session with greetings and end with good byes.
c. **Conversational topic**: The group explicitly opened and closed the topic analyzed in this case study.
d. **Discourse move**: Each move was executed with a single adjacency pair.
e. **Adjacency pair**: An adjacency pair includes an elicitation and a response by each participant.
f. **Textual utterance**: A text chat posting is defined by a participant actively typing and sending a message.
g. **Indexical reference**: A reference is made by a word choice or graphical action by a participant.

This case study provides an unusually clear and simple example of problem solving in a virtual math team. In earlier sessions, the students encountered many difficulties, although they also achieved a variety of successes and learned much about both collaboration and mathematics. At the beginning of their first session, they did not know how to behave together and showed rather poor collaboration skills. Bwang said very little in English, often simply producing drawings or mathematical expressions. Aznx, at the other extreme, tried hard to engage the others, but seemed to have a weak mathematical understanding of what the others were discussing. At various points in the sessions, misunderstandings caused major detours and breakdowns in the group work. Moreover, from an analyst's perspective the interaction was often almost impossible to parse (Stahl, 2009b). By contrast, in the final conversational topic, which is here reviewed, the interaction is focused on two participants; they work well together; they seem to follow each other well; and their work goes quite smoothly. The structure of the interaction is also relatively easy to follow.

It seems that Aznx and Bwang have substantially increased their skills in online collaborative mathematics. The level of their excitement—especially in the segment of Log 8—shows they are highly motivated. Log 10 indicates that they would like to continue this kind of experience in the future.

The Structure of Group Cognition

The analysis of the case study in this paper provides a first analysis of the long-sequence-of-moves structure of collaborative mathematical problem solving in a virtual math team. This is a paradigmatic example of group cognition. The small group—here reduced to a dyad—solves a math problem whose solution had until then eluded them (and had escaped Team C as well).

The students accomplish the problem solving by successively completing a sequence of ten moves. Each of the moves seems almost trivial, but each takes place through an interaction that involves both students in its achievement. The moves are commonplace, taken-for-granted practices of mathematical problem solving. They are familiar from individual and classroom problem solving in algebra classrooms. They have also been encountered repeatedly by Team B in their previous four hours of collaborative problem solving (Medina, Suthers & Vatrapu, 2009). It is this sequence of moves that accomplishes the problem solving. The sequence has an inner logic, with each move requiring the previous moves to have already been successfully completed and each move preparing the way for the following ones.

The common assumption about mathematical problem solving is that math facts and manipulations are what are most important. In our analysis of problem solving in a group context, math content is simply, unproblematically included in individual postings. In fact, more often than not, it is implicitly used and understood "between the lines" of the text chat. Of course, this is only possible because the group had already co-constructed a 'joint problem space' (Kershner et al., 2010; Medina, Suthers & Vatrapu, 2009; Sarmiento & Stahl, 2008; Teasley & Roschelle, 1993) that included this math content as already meaningful for the group. Rather, the important aspects of discourse engaged in collaborative math problem solving are matters of coordination, communication, explanation, decision making and perspective shifting (e.g., moving between visual, verbal and symbolic modes (Çakır, Zemel & Stahl, 2009)). To some extent, these are interactional moves required by most group activities; to some extent, these are adapted to the nature of mathematical discourse.

In conclusion, the group-cognitive achievement of the solution to the group's final problem was accomplished by a sequence of moves. Each move was mundane when considered by itself. The moves and their sequencing were common practices of mathematical problem solving. Each move was interactively achieved through the exchange of base adjacency pairs situated in the ongoing discourse. The problem solving was an act of group cognition structured as a sequence of these interactive moves.

Perhaps this case study can serve as an unusually clear and simple model of the structure of group cognition in mathematical problem solving by a virtual math team. It shows the group cognition taking place through the co-construction of a temporal sequence of problem-solving moves in the group discourse. While the fine structure adheres to the adjacency-pair system of interactional exchange, the larger problem-solving structure builds on these elements through a sequence defined by the topical moves of mathematical deduction.

References

Çakir, M. P., Xhafa, F., & Zhou, N. (2009). Thread-based analysis of patterns in VMT. In G. Stahl (Ed.), *Studying virtual math teams.* (ch. 20, pp. 359-371). New York, NY: Springer. Web: http://GerryStahl.net/vmt/book/20.pdf Doi: http://dx.doi.org/10.1007/978-1-4419-0228-3_20

Çakır, M. P., Zemel, A., & Stahl, G. (2009). The joint organization of interaction within a multimodal CSCL medium. *International Journal of Computer-Supported Collaborative Learning. 4*(2), 115-149. Web: http://GerryStahl.net/pub/ijCSCL_4_2_1.pdf Doi: http://dx.doi.org/10.1007/s11412-009-9061-0

Kershner, R., Mercer, N., Warwick, P., & Staarman, J. K. (2010). Can the interactive whiteboard support young children's collaborative communication and thinking in classroom science activities? *International Journal of Computer-Supported Collaborative Learning. 5*(4)

Medina, R., Suthers, D. D., & Vatrapu, R. (2009). Representational practices in VMT. In G. Stahl (Ed.), *Studying virtual math teams.* (ch. 10, pp. 185-205). New York, NY: Springer. Web: http://GerryStahl.net/vmt/book/10.pdf Doi: http://dx.doi.org/10.1007/978-1-4419-0228-3_10

Sacks, H. (1962/1995). *Lectures on conversation*. Oxford, UK: Blackwell.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language. 50*(4), 696-735. Web: www.jstor.org

Sarmiento, J., & Stahl, G. (2008). *Extending the joint problem space: Time and sequence as essential features of knowledge building.* Paper presented at the International Conference of the Learning Sciences (ICLS 2008). Utrecht, Netherlands. Web: http://GerryStahl.net/pub/icls2008johann.pdf

Schegloff, E., & Sacks, H. (1973). Opening up closings. *Semiotica. 8*, 289-327.

Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge, UK: Cambridge University Press.

Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge*. Cambridge, MA: MIT Press. 510 + viii pages. Web: http://GerryStahl.net/mit/

Stahl, G. (2009a). *Keynote: How I view learning and thinking in CSCL groups.* Paper presented at the International Conference on Computers and Education (ICCE 2009). Hong Kong, China. Web: http://GerryStahl.net/pub/icce2009keynote.pdf

Stahl, G. (2009b). Meaning making in VMT. In G. Stahl (Ed.), *Studying virtual math teams.* (ch. 26, pp. 505-527). New York, NY: Springer. Web: http://GerryStahl.net/vmt/book/26.pdf Doi: http://dx.doi.org/10.1007/978-1-4419-0228-3_26

Stahl, G. (2009c). *Studying virtual math teams*. New York, NY: Springer. 626 +xxi pages. Web: http://GerryStahl.net/vmt/book Doi: http://dx.doi.org/10.1007/978-1-4419-0228-3

Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences.* (pp. 409-426). Cambridge, UK: Cambridge University Press. Web: http://GerryStahl.net/cscl/CSCL_English.pdf

Stahl, G., Zemel, A., & Koschmann, T. (2009). *Repairing indexicality in virtual math teams.* Paper presented at the International Conference on Computers and Education (ICCE 2009). Hong Kong, China. Web: http://GerryStahl.net/pub/icce2009.pdf

Suthers, D., Dwyer, N., Medina, R., & Vatrapu, R. (2010). A framework for conceptualizing, representing, and analyzing distributed interaction. *International Journal of Computer-Supported Collaborative Learning. 5*(1), 5-42. Web: http://dx.doi.org/10.1007/s11412-009-9081-9

Teasley, S. D., & Roschelle, J. (1993). Constructing a joint problem space: The computer as a tool for sharing knowledge. In S. P. Lajoie & S. J. Derry (Eds.), *Computers as cognitive tools.* (pp. 229-258). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Zemel, A., & Çakir, M. P. (2009). Reading's work in VMT. In G. Stahl (Ed.), *Studying virtual math teams.* (ch. 14, pp. 261-276). New York, NY: Springer. Web: http://GerryStahl.net/vmt/book/14.pdf Doi: http://dx.doi.org/10.1007/978-1-4419-0228-3_14

Zemel, A., Xhafa, F., & Çakir, M. P. (2009). Combining coding and conversation analysis of VMT chats. In G. Stahl (Ed.), *Studying virtual math teams.* (ch. 23, pp. 421-450). New York, NY: Springer. Web: http://GerryStahl.net/vmt/book/23.pdf Doi: http://dx.doi.org/10.1007/978-1-4419-0228-3_23

Constructing Part-task Congruent Representations to Support Coordination of Collaborative Problem-solving Tasks

B. Slof, University of Groningen, 9700 AV Groningen, The Netherlands, B.Slof@rug.nl
G. Erkens, Utrecht University, 3508 TC Utrecht, The Netherlands, G.Erkens@uu.nl
P. A. Kirschner, Open University, 6401 DL Heerlen, The Netherlands, P.A.Kirschner@ou.nl

Abstract: This study investigated whether constructing part-task congruent domain-specific representations supports teams in establishing and maintaining a shared understanding of the knowledge domain and negotiating about it. This better coordination of team' discussions was, in turn, expected to lead to better problem-solving performance. In triads secondary education students worked on a case-based business-economics problem in a predefined order, but differed in the representational tool(s) they received. In the matched condition, teams received the three tools in a part-task congruent manner. In the other three non-matched conditions, teams received one of the tools for all three part-tasks, thus a tool congruent to one part-task and incongruent for the other two. The results show that coordination processes were indeed better which might explain why teams in the matched condition performed better on the complex problem-solving task. However, similar results were obtained by teams who only received a tool for constructing causal representations for all part-tasks.

Introduction

There has been a recent surge in the interest of educational researchers for studying the effects of computer-supported tools for fostering team' complex learning-task performance (e.g., Slof, Erkens, Kirschner, Jaspers, & Janssen, 2010). Carrying out complex learning tasks, such as solving complex problems, often requires learners to actively engage in a dynamic process of sense-making by articulating and discussing multiple representations on the problem and their problem-solving strategy. Through externalizing one's knowledge, discussing this with peers, and establishing and refining the teams' shared understanding of the domain, learners may acquire new knowledge and skills and process them more deeply (e.g., Hmelo-Silver, Duncan, & Chinn, 2007). It, however, seems that such meaningful discussions about the domain can hardly be reached when learners are not aware of each others' knowledge, ideas and do not negotiate about them with their peers. In this respect, it is often advocated that collaborative learning situations require three main processes of coordination: (1) mutual activation and sharing of knowledge and skills, (2) grounding or creating a common frame of reference, and (3) negotiation or the process of coming to agreement (e.g., Erkens & Janssen, 2008). Research on Computer Supported Collaborative Learning (CSCL) has shown that collaboratively constructing and discussing external representations can beneficially affect complex learning-task performance. Embedding representational tools in a CSCL-environment can facilitate learners' construction of different representations of the domain through its representational guidance and, thereby, guide their domain-specific interaction (Fischer, Bruhn, Gräsel, & Mandl, 2002; Suthers, 2006). Furthermore, in their discussions learners can refer to the constructed representation (i.e., deictic referencing), thereby supporting them to create a common frame of reference and facilitating a meaningful discussion (Suthers, Hundhausen, & Girardeau, 2003). These studies, though very valuable, often neglect the fact that problem-solving tasks are usually composed of fundamentally different phase-related part-tasks (1) problem-orientation (i.e., determining core concepts and relating them to the problem), (2) problem-solution (i.e., proposing solutions to the problem) and (3) solution-evaluation (i.e., determining suitability of the solutions and coming to a definitive solution to the problem). Important here is that each part-task mostly requires a different domain-specific representation and, thus, requires a tool with a different kind of representational guidance. When the design of the tool is incongruent with the demands of one or more part-tasks, learners may experience communication problems and problem-solving performance might decrease (e.g., Van Bruggen, Boshuizen, & Kirschner, 2003). The study presented in this paper is aimed at determining whether (1) proper coordination process can be evoked through constructing part-task congruent representations and (2) such an approach can lead to better problem-solving performance.

Coordination Processes

For meaningful discussion to arise, learners have to properly coordinate their discussions of the concepts, principles and procedures by carrying out communicative activities such as (1) making their own knowledge and ideas explicit to other group members, (2) focusing, (3) checking and (4) argumentation (Andriessen, Baker, & Suthers, 2003; Erkens & Janssen, 2008). When made explicit, learners have to try to maintain a shared topic of discourse and to repair a common focus if they notice a focus divergence. Learners coordinate their topic of discourse by *focusing*. Also, not all concepts, principles, and procedures are relevant for carrying out a part-task, thus, learners also have to guard the coherence and consistency of their shared understanding of the knowledge domain (e.g., Erkens & Janssen, 2008). By *checking*, learners ground their communication in a common understanding, which was found to be one of the major communicative activities in dialogues of collaborative problem-solving and related to the quality of the problem solving process (e.g., Van der Linden, Erkens, Schmidt, & Renshaw, 2000). Furthermore, learners have to come to agreement with respect to relevant concepts principles and procedures. By *argumentation* they will try to change their partners viewpoint to arrive at the best way to carry out a part-task or at a definition of concepts acceptable for all. In this argumentation process they try to convince the other(s) by elaborating on their point of view, giving explanations, justifications and accounts (e.g., Kirschner, Beers, Boshuizen, & Gijselaers, 2008).

Constructing Part-task Congruent Representations

In the research reported on in this article, learners collaborated on solving a case-based business-economics problem in which they had to advise an entrepreneur about changing the business strategy to increase profits. Due to its different part-tasks, such a problem-solving task require multiple representational tools facilitating the construction of different representations. The specific ontology (i.e., objects, relations, rules for combining them) of each tool guides learner interaction in a specific manner by supporting them in using restricted views of the knowledge domain (i.e., problem representations). To effectively do this, one must carefully match the tools' ontology to the different part-task demands (Van Bruggen et al., 2003; Slof et al., 2010). Scripting was employed to ensure this alignment between tool, tool use and part-task demands. According to Dillenbourg a script is "a set of instructions regarding to how the group members should interact, how they should collaborate and how they should solve the problem" (2002, p. 64). Integrating scripting with the availability of representational tools sequences and makes the different part-task demands explicit so that they can be foreseen with part-task congruent guidance in the representational tools (see Table 1). By doing so, communicative activities beneficial for coordinating the collaborative problem-solving might be evoked.

Table 1: Congruence between representational tool and phase-related part-task demands.

Problem phase	Task demands	Representational tool	Representational guidance
Problem-orientation	Determining core concepts and relating them to the problem	Conceptual	Visualizing concepts and their conceptual relationships
Problem-solution	Proposing multiple solutions to the problem	Causal	Visualizing causal relationships between the concepts and the possible solutions
Solution-evaluation	Determining suitability of the solutions and coming to a definitive solution to the problem	Simulation	Visualizing mathematical relationships between the concepts and enabling manipulation of their values

In the *problem-orientation* phase learners have to explain what they think the problem is and describe what the most important factors are for solving it. The interaction should, therefore, be guided towards selecting the core concepts needed to carry out this part-task and discussing how those concepts are qualitatively related to each other. The design of the representational tool should facilitate learners in constructing and discussing a global qualitative problem representation by guiding and supporting them in conceptually relating the relevant concepts. Figure 1 shows an experts' representation of the concepts and their conceptual interrelationships involved in this study. The conceptual representational tool facilitates representation of the concepts and their interrelationships shown in Figure 1. Selecting and relating concepts that the learners may regard as beneficial for solving the problem supports them in becoming more familiar with those concepts and in broadening their problem space. Learners receiving the conceptual tool could, for example, make explicit that the 'company result' is related to the 'total profit' and 'efficiency result'. This should guide those learners in elaborating (i.e., causal, mathematical) on the relationships in the two following problem phases, making it easier for them to find multiple solutions to the problem and to evaluate their effects.

In the *problem-solution* phase learners have to formulate several changes of the business strategy (i.e., interventions) and make clear how they might solve the problem (i.e., problem-solution) by describing how they will affect the outcomes (i.e., company results). The interaction should, thus, be guided towards formulating multiple interventions and discussing how each of these interventions affects the selected core concepts by further specifying the relationships between the concepts and the proposed interventions. The representational tool should facilitate construction and discussion of a causal problem representation by causally relating concepts to each other and to possible interventions. Figure 2 shows an experts' representation of the concepts, the possible interventions and their causal interrelationships involved in this study. The causal representational tool facilitates representation of the concepts, interventions and their interrelationships shown in Figure 2. Selecting relevant concepts and interventions and causally relating them supports the effective exploration of the solution space and, thus, of finding multiple solutions to the problem. Learners receiving the causal representational tool could, for example, make explicit that an intervention such as a employing a promotion-campaign (e.g., placing an advertisement in a paper) affects 'actual sales', which in turn affects 'total profit'. Only conceptually representing the interrelationships of the concepts, as in the first problem phase, is not expressive enough for this part-task since the relationships need to be further specified and learners need additional information about the possible solutions. If this is not the case, then learners are forced to come up with a solution (i.e., the advice) themselves without sufficient understanding of the underlying qualitative principles governing the domain.

Finally, in the *solution-evaluation phase* learners have to determine the financial consequences of their proposed interventions and formulate a suitable and definitive advice for the entrepreneur by discussing the suitability of the different interventions with each other. The interaction should, therefore, be guided towards determining and comparing the financial consequences by discussing the mathematical relationships between the selected concepts. The representational tool must, thus, facilitate constructing and discussing a quantitative representation by specifying the relationships as equations. Figure 3 shows an experts' representation of the concepts and their mathematical interrelationships involved in this study. The simulation representational tool facilitates representation of the concepts and their interrelationships shown in Figure 3. Selecting relevant concepts and specifying the interrelationships as equations supports learners in evaluating the effects of their proposed interventions and, thus, in coming to a suitable advice. Learners receiving the simulation representational tool could, for example, simulate how an intervention such as employing a promotion-campaign affects the 'actual sales' and whether this affects the 'total profit'. By entering values and adjusting them (i.e., increasing or decreasing), the values of the other related concepts are automatically computed. Since such quantitative representations can only be properly understood and applied when learners have well-developed qualitative understanding of the domain, this kind of support is only appropriate for carrying out this type of part-task.

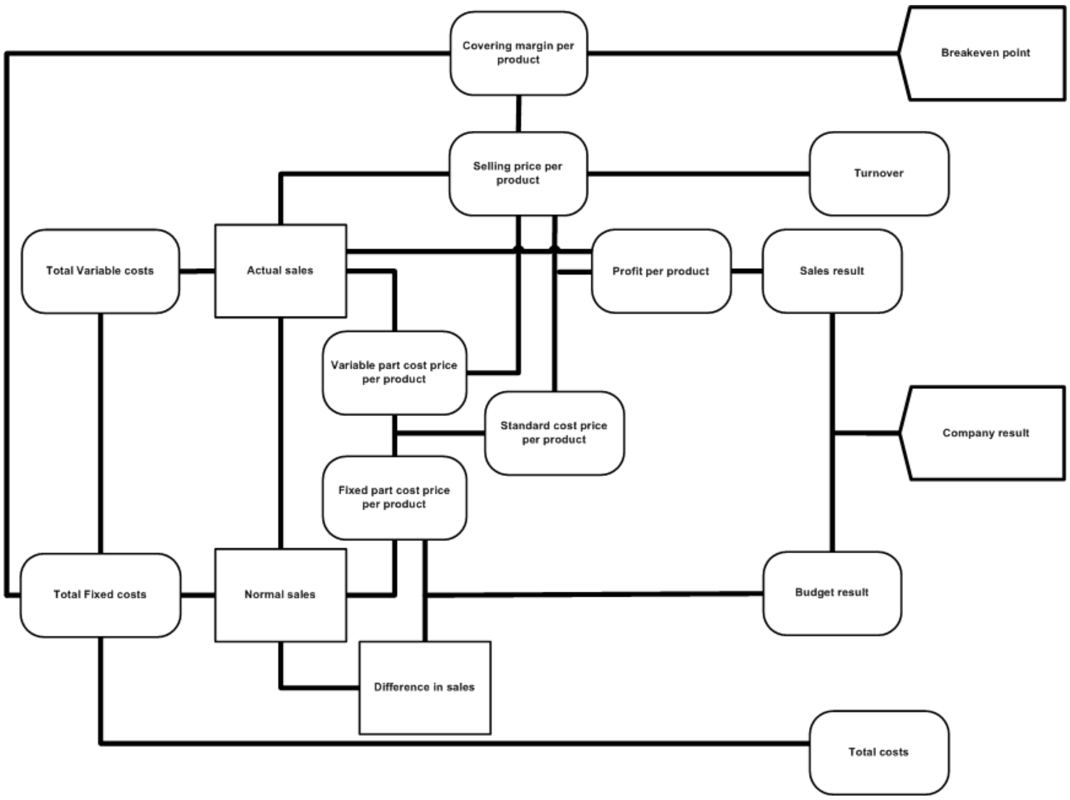


Figure 1. Experts' Conceptual Representation of the Domain.

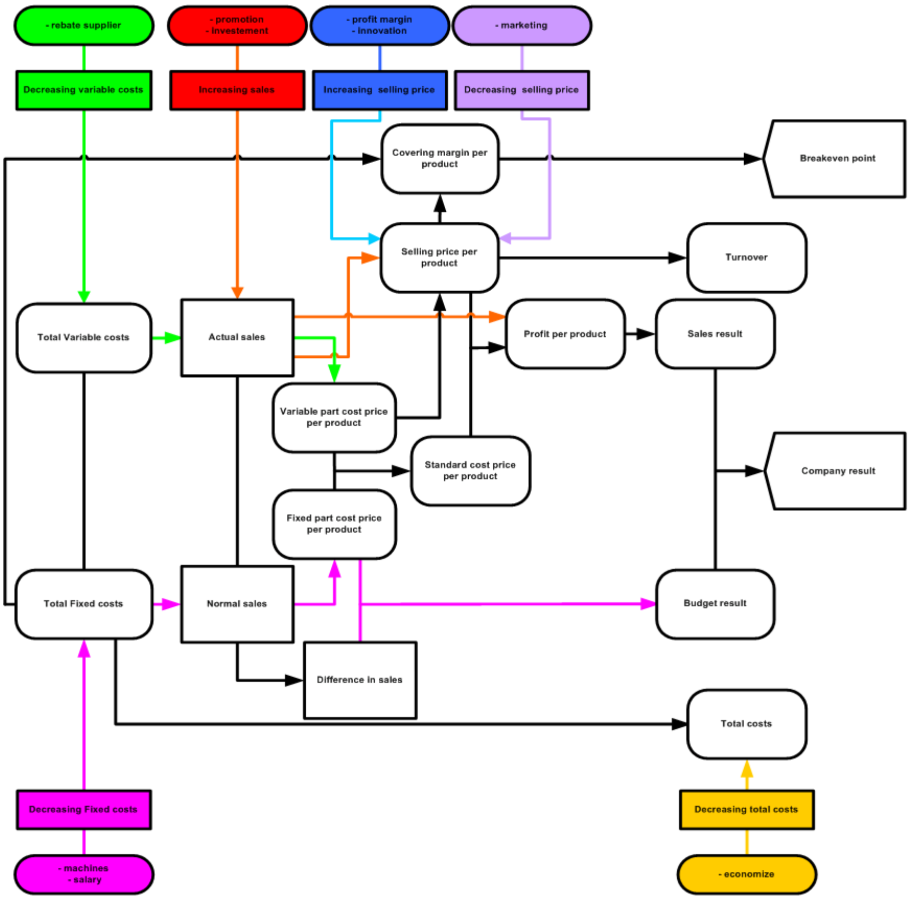


Figure 2. Experts' Causal Representation of the Domain.

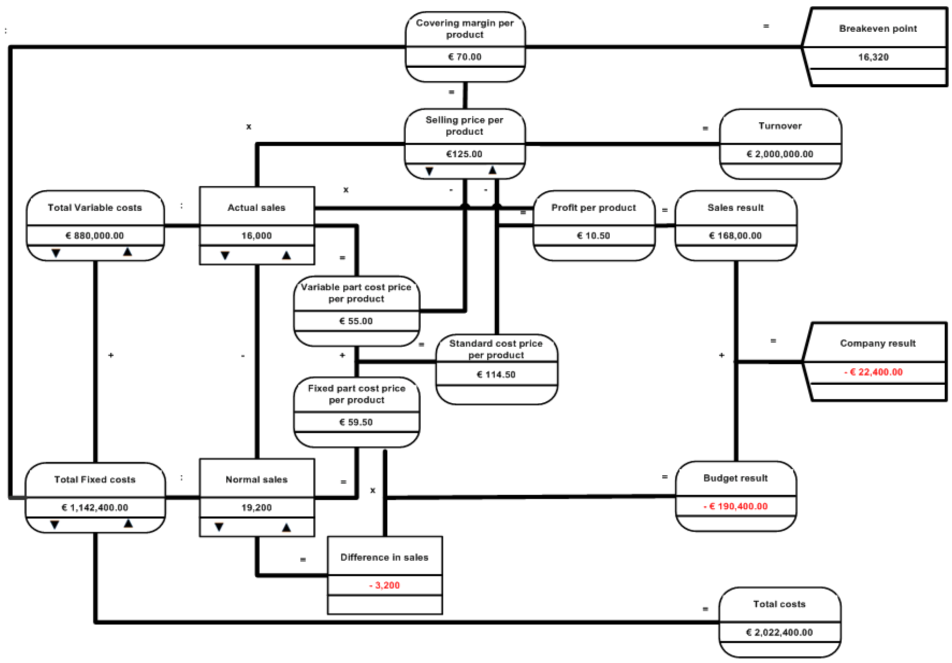


Figure 3. Experts' Mathematical Representation of the Domain.

Design and Expectations

The research reported on here was aimed at determining whether constructing part-task congruent representations affects both teams' communicative activities and problem-solving performance in a CSCL-environment. In four experimental conditions, student triads had to collaboratively solve a case-based problem in business-economics that was divided into three problem phases each coupled with a different representational tool. To study the effect of condition, the tools' representational guidance was either matched or mismatched to the different problem phases; in other words it was either congruent or incongruent with the required task activities (see Table 2).

Table 2: Overview of the experimental conditions.

Problem phase	Condition and provided representational tool			
	Conceptual condition	Causal condition	Simulation condition	Matched condition
Problem-orientation	*Conceptual tool*	Causal tool	Simulation tool	*Conceptual tool*
Problem-solution	Conceptual tool	*Causal tool*	Simulation tool	*Causal tool*
Solution-evaluation	Conceptual tool	Causal tool	*Simulation tool*	*Simulation tool*

Teams in all conditions were scripted to carry out all the part-tasks in a predefined order, but differed in the representational tool(s) - conceptual, causal or simulation - they received. In the *matched condition*, teams received the three tools in a part-task congruent manner. In the other three *non-matched conditions*, teams received one of the tools for all three part-tasks, thus a tool congruent to one part-task and incongruent for the other two. Due to this presumed match between tools' representational guidance and the part-tasks, it was hypothesized that teams in the matched condition would *(H1)* experience a qualitatively better learning process, evidenced by carrying out more communicative activities to coordinate their collaborative problem-solving process and *(H2)* achieve a better problem-solving performance, evidenced by arriving at better interventions.

Method

Participants

Participants were students from six business-economics classes in three secondary education schools in the Netherlands. The total sample consisted of 93 learners (60 male, 33 female; mean age = 16.74 years; SD = .77, Min = 15, Max = 18). The students were, within classes, randomly assigned to a total of 31 teams of learners (i.e., triads); seven teams in the matched condition and eight teams in each of the three non-matched conditions.

Problem-solving Task and Materials

CSCL-environment: Virtual Collaborative Research Institute

Teams worked in a CSCL-environment called Virtual Collaborative Research Institute (VCRI, see Figure 4; Jaspers, Broeken, & Erkens, 2005).

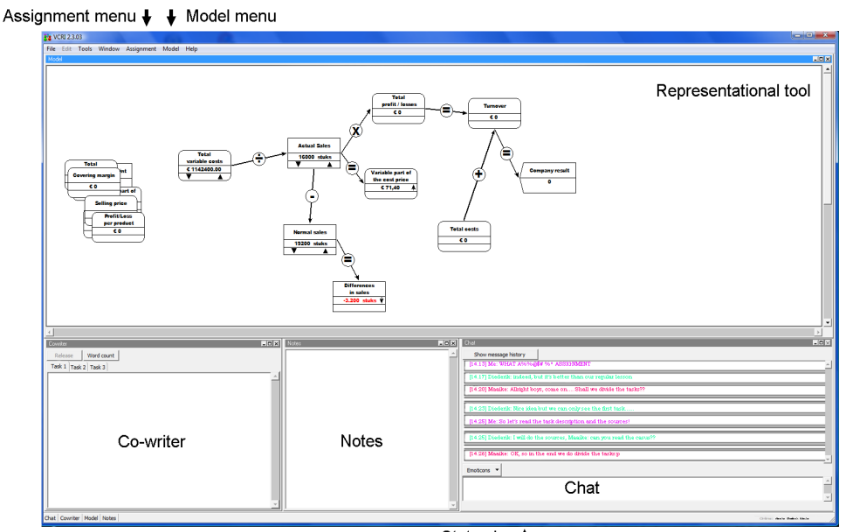


Figure 4. Screenshot of the VCRI-program; Simulation Representational Tool.

Problem-solving Task and Part-task Congruent Representations

All teams were coerced to carry out the part-tasks in a predefined order (i.e., used the same script) and could, thus, only start with a new part-task after finishing an earlier part-task. When team members agreed that a part-task was completed, they had to 'close' that phase in the assignment menu. This 'opened' a new phase, which had two consequences for all learners, namely they were instructed to (1) carry out a new part-task and (2) revise their representation of the domain so it concurred with the decisions they gave to the new part-task. Learners in the non-matched conditions were facilitated in elaborating on their previously constructed representation. Since those learners kept the same representational tool, all concepts and their relationships remained visible and could be revised as learners seemed appropriate for carrying out their new part-task. Learners in the matched condition were facilitated in using a different qualitative or quantitative perspective of the domain. That is, the previously selected concepts remained visible and learners were instructed to replace the relationships by specifying them in a causal manner or as equations with the aid of their new representational tool.

Procedure

All teams spent six, 45-minute, lessons solving the problem during which each team member worked on a separate computer. Before the first lesson, learners received an instruction about the CSCL-environment, the complex learning task, and the team composition. students worked on the problem in the computer classroom where all actions and decisions were logged.

Measures

Coordination Processes

To examine the effect of condition data concerning learners' coordination processes was collected by logging the chat-utterances of the group members. A dialogue act is regarded as a communicative action that is elicited for a specific purpose representing a specific function in the dialogue (Erkens & Janssen, 2008). Dialog-act coding was based on the occurrence of characteristic words or phrases (i.e., discourse markers) that indicated the communicative function of an utterance. The chat-protocols were searched for the occurrence of these discourse markers that led to the identification and coding of the dependent variables (see Table 4). This was automatically done with a MEPA-filter using 'if-then' decision rules that uses pattern matching to find typical words or phrases. Reliability of the dialogue act coding filter compared to hand-coding is 79% (Erkens & Janssen, 2008). After coding, score-frequencies for each dialogue act were computed and combined resulting in the dependent variables.

Table 3: Coding of learners' communicative activities.

Activities	Dialogue Act	Description	Example discourse marker
Focusing	Elicitative proposal for action	Proposition for action	Shall we get started with the first part-task?
	Elicitative question open	Open question with a lot of alternatives	What do you think we should do next?
	Imperative action	Command to perform an action	Write the conclusion
	Imperative focus	Command for attention	Look at the representational tool
	Elicitative question verify	Question that can only be answered with yes or no	Do you refer to the company result?
Checking	Elicitative question set	Question where the alternatives are already given (set)	Are you for / against increasing sales?
	Responsive confirm	Confirming answer	Yes, sure
	Responsive deny	Denying answer	No, not
	Responsive accept	Accepting answer	Oh, Yes
Argumentation	Argumentative reason	Reason	Because….
	Argumentative against	Objection	But …
	Argumentative conditional	Condition	If … (then …)
	Argumentative then	Consequence	Then …
	Argumentative disjunctive	Disjunctive	…or… or
	Argumentative conclusion	Conclusion	Thus…

Problem-solving Performance

To measure the effect of condition on problem-solving performance an assessment form for criteria of the problem-solving task, such as correctness, elaborateness and suitability of the decisions to all part-tasks, was developed. All items were coded as; 0, 1 or 2, whereby a '2' was coded when the answer given was of high quality (e.g., more suitable).

Data Analyses

In CSCL, team members influence each other (i.e., behave more or less similarly) causing non-independence of measurement (Kenny, Kashy, & Cook, 2006). This is problematic because many statistical techniques assume score independence and a violation compromises interpretation of the analyses. Non-independence was determined by computing the intraclass correlation coefficient and its significance (Kenny, et al., 2006), for all dependent variables for student interaction. This resulted in non-independence (α < .05) for all tests, justifying *Multilevel analysis* (MLA). MLA compares the deviance of an empty model and a model with one or more predictor variables to compute a possible decrease in deviance. The latter model is considered a better model when there is a significant decrease in deviance compared to the empty model (tested with a χ^2-test).

Results

Coordination Processes

MLAs revealed that condition was a (marginally) significant predictor for the communicative activities learners exhibited when comparing learners in the matched condition to those in both the conceptual (β = 23.84, *p* = .06) and the simulation conditions (β = 42.00, *p* = .00). When analyzing learners' communicative activities for the conditions separately, several category effects were found (see Table 4). First, a significant category effect for *focusing* was found; learners in the matched condition were better able to coordinate what their topic of discourse was than learners in both the conceptual (β = 4.22, *p* = .05) and the simulation conditions (β = 6.68, *p* = .02). Second, a significant category effect for *checking* was found; learners in the matched condition devoted more attention to guarding the coherence and consistency of their shared understanding of the domain than learners in both the conceptual (β = 14.08, *p* = .04) and the simulation conditions (β = 23.03, *p* = .00). Finally, a significant category effect was found for *argumentation*; learners in the matched condition exhibited more argumentative activities than learners in the simulation condition (β = 12.17, *p* = .02). As expected, learners in the matched condition were better able to establish and maintain shared understanding of the domain than learners in both the conceptual and simulation conditions. These differences were, however, not obtained for the comparison with learners in the causal condition.

Table 4: Multilevel analyses for effects of condition concerning communicative activities.

	Conceptual condition ($n_{learner}$ = 24)	Causal condition ($n_{learner}$ = 24)	Simulation condition ($n_{learner}$ = 24)	Matched condition ($n_{learner}$ = 21)	Effects matched condition $N_{learner}$ = 93)		
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	$\chi^2(3)$	β	*SE*
Coordination *	124.33 (59.01)	173.82 (130.42)	87.65 (54.21) -	170.36 (79.22) +	30.06	24.02	19.40
Focusing *	22.87 (8.20) -	31.50 (23.37)	18.13 (12.28) -	31.09 (15.83) +	18.42	4.30	3.52
Checking *	57.33 (31.43) -	88.95 (69.43)	39.17 (26.471) -	84.14 (38.56) +	27.74	14.22	9.84
Argumentation *	44.12 (26.92)	53.36 (43.65)	30.35 (19.95) -	55.14 (32.18) +	20.90	5.41	6.61

Note. * *p < .05,* ** *p < .01*; if matched condition significantly > a non-matched condition than the matched condition is indicated with a + and the non-matched condition with a -.

Problem-solving Performance

A one-way MANOVA on the total score for teams' problem-solving performance showed a significant difference for condition (*F*(3,27) = 4.38, p = .01). Bonferroni post hoc analyses revealed that teams in the matched condition scored significantly higher than teams of learners in both the conceptual (*p* = .01; *d* = 1.46) and the simulation conditions (*p* = .01; *d* = 1.48). When the results for the dependent variables were considered separately - using one-way ANOVAs with Bonferroni post hoc analyses - condition effects were found for 'justification' (*F*(3,27) = 4.85, *p* = .01) and 'correctness' (*F*(3,27) = 3.97, *p* = .01). First, teams in the matched condition provided more arguments for the decisions to the part-tasks (i.e., justification) than teams in both the conceptual (*p* = .01; *d* = 1.56) and simulation conditions (*p* = .01; *d* = 1.56). Second, teams in the matched condition used the business-economics concepts and their interrelationships more correctly in their decisions to the different part-tasks (i.e., correctness) than teams in both the conceptual (*p* = .01; *d* = 3.97) and simulation conditions (*p* = .03; *d* = 2.52). As expected, teams of learners receiving part-task congruent representational tools scored higher on problem-solving performance. Although expected, no significant differences were obtained between teams in the matched condition and the causal condition.

Discussion

The results indicate that teams of learners that received the complete array of tools (i.e., matched condition) were indeed fostered in their problem-solving performance. That is, those teams formulated better decisions with respect to the part-tasks and came up with better definitive solutions to the problem than learners in both the conceptual and the simulation conditions. This might be explained by the fact that learners in the matched condition exhibited more communicative activities than those in both the conceptual and simulation conditions. That is, they were better able to establish and maintain a shared understanding of the domain, which is regarded as a prerequisite for having a meaningful discussion of the domain (e.g., Van der Linden et al., 2000). It seems that the deictic power of the representational tool hindered learners in establishing and maintaining shared understanding of the domain (Suthers et al., 2003). In other words, when learners are unable to specify (i.e., conceptual tool) or being forced to explicitly specify (i.e., simulation tool) the relationship between concepts this hinders learners in properly referring to and relating their contributions in CSCL-environments. Although the results seem very promising, problem-solving performance of teams in the causal condition was very similar to what was found in the matched condition. Since teams in both conditions received the causal representational tool they were both facilitated in constructing and discussing a causal domain representation. Supporting learners' causal reasoning seems, thus, important for problem-solving. This result raises questions about whether constructing and applying multiple representations of a domain is beneficial for problem-solving performance. Additional research seems, therefore, needed to investigate whether learners:

- require qualitative as well as quantitative representations during their complex learning-task performance,
- combine qualitative and quantitative representations during their complex learning-task performance.

References

Andriessen, J., Baker, M., & Suthers, D. D. (2003). Argumentation, computer support, and the educational context of confronting cognitions. In J. Andriessen, M. Baker, & D. D. Suthers, (Eds.), *Cognitions: Arguing to learn* (pp. 1-25). Dordrecht, The Netherlands: Kluwer Academic Press.

Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. In P. A. Kirschner (Ed.), *Three worlds of CSCL: Can we support CSCL?* (pp. 61-91). Heerlen, The Netherlands: Open Universiteit Nederland.

Erkens, G., & Janssen, J. (2008). Automatic coding of online collaboration protocols. *International Journal of Computer-Supported Collaborative Learning, 3,* 447–470.

Fischer, F., Bruhn, J., Gräsel, C., & Mandl, H. (2002). Fostering collaborative knowledge construction with visualization tools. *Learning and Instruction, 12,* 213–232.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42,* 99–107.

Jaspers, J. G. M., Broeken, M., & Erkens, G. (2005). Virtual Collaborative Research Institute (VCRI). Version 2.2. Utrecht, The Netherlands: Utrecht University.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York/London: The Guilford Press.

Kirschner, P. A., Beers, P. J., Boshuizen, H. P. A., & Gijselaers, W. H. (2008). Coercing shared knowledge in collaborative learning environments. *Computers in Human Behavior*, *24*, 403–420.

Slof, B., Erkens, G., Kirschner, P. A., Jaspers, J. G. M., & Janssen, J. (2010). Guiding students' online complex learning-task behavior through representational scripting. *Computers in Human Behavior, 26,* 927–939.

Suthers, D. D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. *International Journal of Computer-Supported Collaborative Learning,1,* 315–337.

Suthers, D. D., Girardeau, L., & Hundhausen, C. (2003). Deictic Roles of External Representations in Face-to-face and Online Collaboration. Designing for Change in Networked Learning Environments. In B. Wasson, S. Ludvigsen & U. Hoppe (Eds), *Proc. International Conference on Computer Support for Collaborative Learning* (pp. 173-182). Dordrecht: Kluwer Academic Publishers.

Van Bruggen, J. M., Boshuizen, H. P. A., & Kirschner, P. A. (2003). A cognitive framework for cooperative problem solving with argument visualization. In P. A. Kirschner, S. J. Buckingham Shum, & C. S. Carr (Eds.), *Visualizing Argumentation: Software tools for collaborative and educational sense-making.* (pp. 25-47). London: Springer.

Van der Linden, J. L., Erkens, G., Schmidt, H., & Renshaw, P. (2000). Collaborative learning. In P. R. J. Simons, J. L. Van der Linden, & T. Duffy (Eds.), *New learning* (pp. 1-19). Dordrecht: Kluwer Academic Publishers.

Virtual Mathematical Inquiry: Problem Solving at the Gestural–Symbolic Interface of Remote-Control Embodied-Interaction Design

Dragan Trninic, José F. Gutiérrez, Dor Abrahamson, Embodied Design Research Laboratory, Graduate School of Education, University of California at Berkeley, CA 94720-1670 USA
Email: trninic@berkeley.edu, josefrancisco@berkeley.edu, dor@berkeley.edu

Abstract: What, if any, are unique affordances of embodied-interaction (EI) design for mathematics learning? How, in turn, might engineering and testing such design aid us in building theory of learning? We draw on media studies to argue that activities situated at the hybrid intersection of two media, for example physical-virtual interfaces, enable both learners and their researchers to objectify important notions. We evoke cultural-historical psychology theory, which implicates gesture as the ontogenesis of inscription, to interpret instructional EI as offering analogous enactive-semiotic tension and resolution. We support our view with selected empirical data from a design-based research project investigating the emergence of mathematical concepts from mediated engagement in EI activity. Specifically, we present, analyze, and juxtapose two case studies respectively illustrating student success or failure to grope toward a conceptual system (proportions) embedded in a technological system (EI).

> The hybrid or the meeting of two media is a moment of truth and revelation from which new form is born….The moment of the meeting of media is a moment of freedom and release from the ordinary trance and numbness imposed….on our senses. (McLuhan, 1964, p. 63)

Objective: The Unique Affordances of Embodied-Interaction Pedagogical Design as Context for Research on Mathematics Learning

This is a theory-meets-design paper. The design in question is *embodied interaction* (EI), a form of technology-supported multimodal training activity. Through engaging in EI activities, users are expected to build schematic perceptuomotor structures consisting of mental connections between, on the one hand, physical actions they perform as they attempt to solve problems or respond to cues and, on the other hand, automated sensory feedback on these actions. Emblematic of EI activities, and what distinguishes EI from "hands on" educational activities in general, whether involving concrete or virtual objects, is that EI users' physical actions are intrinsic, and not just logistically instrumental, to obtaining information (cf. Marshall, Cheng, & Luckin, 2010). That is, the learner is to some degree physically immersed in the microworld, so that finger, limb, torso, or even whole-body movements are not only in the service of acting upon objects but rather the motions themselves become part of the perceptuomotor structures learned. EI is not simply "hands on" but "hands *in*." The theoretical issue we seek to address here is how the rising cognitive-science theory of embodied cognition (e.g., Barsalou, 2008; Hommel, Müsseler, Aschersleben, & Prinz, 2001) and related theoretical debates over the epistemic nature of enaction might influence design-based researchers' conceptualization and implementation of EI designs. We contextualize this exploration in our recent efforts to create and research EI activities for mathematics learning.

As a point of departure, consider the following quotation from a recent *Interacting with Computers* special issue on "Enactive Interfaces," where Chris Raymaekers characterizes research into enaction as "[striving] to create new types of interfaces, which make use of a user's capacities to learn how a system works by using it… The computer is in this case a tool that is as invisible to the user as possible: the user interacts through the computer" (2009, p. 2).

In this paper, we—a research team investigating the nature of mathematics learning by developing and implementing innovative instructional technology—reflect on a recent EI study that availed heuristically of this alleged human capacity to "learn how a system works by using it." In a sense, it is this very capacity that we design for, because we attempt to create artifacts that emulate principles of historically successful cultural mediations, such as children learning the numerical system through learning to count. In a recent study, we embedded a targeted mathematical system, proportions, as the normative interaction principle for operating a "mystery" technological device and studied whether and how students' goal-oriented enactment in this EI microworld supports the guided development of commensurate conceptual systems. So doing, however, we encountered certain challenges in enabling students to "learn how a system works by using it." As we will explain, the particular challenges in question here relate to very initial steps of getting "into" our EI media.

When instructional designers create artifacts, students may engage them in unexpected ways that lead them astray of the designers' intended interaction trajectory. This is true of innovative electronic artifacts in general (Olive, 2000), possibly because these artifacts are historically new and so both designers and users are still developing an understanding of their affordances. For example, users of augmented reality technology who manipulate tactile artifacts linked to virtual simulacra tacitly expect the virtual objects to interact in accord with their physical ontology, yet those anticipated affordances may not have been engineered into the interaction space (see Hornecker & Dünser, 2009). EI designs may preempt such false projection by incorporating generic remote controllers or even no controllers at all, such that no assumed ontology is evoked. Yet EI users nevertheless may still encounter challenges in negotiating the physical–virtual gulf. We will attempt, below, to illuminate these challenges of futuristic pedagogical tools by looking back to the works of Vygotksy and McLuhan as well as contemporary philosophers of Human–Computer Interaction (HCI). We thus expand on the CSCL 2011 call by connecting *outside* of the learning sciences to greater circles of inquiry. By looking back to move forward, we hope also to frame the enduring interdependences of educational design and theory.

Background: Learning as Appropriating Artifact-Bound Conceptual Systems

Much of research in education has been divided along a schism between "cognitive" and "sociocultural" perspectives. We posit that this split is a false dichotomy that has, to an extent, undermined our field's efforts in developing and evaluating cogent theories of learning. Thus we are committed to a view of learning and instruction that sees the learner as a resourceful cognitive agent operating in a sociocultural sphere (cf. Karmiloff-Smith, 1988). Ultimately, we concur, "These are not optional perspectives, nor even strictly complementary ones. Each perspective completes the other, and in fact, they are probably not coherently construed independently" (diSessa, 2008, pp. 427-428). By investigating the ontological nature and pedagogical roles of educational artifacts, we hope to gain reliable purchase on the interplay of the cognitive and sociocultural, specifically because our designs are centered on creating opportunities for mediated discovery. In particular, EI design studies should be of interest to scholars who are developing models of learning informed by both cognitive and sociocultural theory, because the studies generate interaction data that bring out in relief mediation practices for the cultural signification of discipline-neutral schematic action. Through investigating EI-based mathematics learning, we are interested in theorizing how social interactions steer participants to leverage perceptuomotor competence in the appropriation of disciplinary forms of reasoning.

The intrinsically embedded, contextual nature of human artifacts, and specifically the ontologically requisite relations between artifacts and their engulfing normative application structures and activity systems, implies that these objects are never "objective." Rather, artifacts might be better understood as the embodiments of certain cultural practices, condensed templates of theory-in-action of doing things as discovered in the history of human civilization (Stetsenko, 2002). As Leont'ev (1981) eloquently writes, the artifact "mediates activity and thus connects humans not only with the world of objects but also with other people. Because of this, humans' activity *assimilates the experience of humankind*" (Leont'ev, 1981, p. 56, original italics).

Thus, adopting a sociocultural view, we espouse a position that learning is the residual effect of interacting with artifacts—or, more precisely, of engaging one's own fluency-delimited mental construction of these artifacts—as means of accomplishing one's goals (cf. Salomon, Perkins, & Globerson, 1991). However, in keeping with our dialectical commitment, we complement this view from a cognitivist perspective to promote a conceptualization of learning also as the iterative modification of existing goal-oriented schemes. That is, against the sociocultural backdrop view of *learning as imitating, internalizing, and appropriating* the elders' artifact-mediated actions, we foreground a view of *learning as discovering* these artifacts' horizons in the course of explorative problem solving, theory building, and creative inference making.

We propose that whereas the discovery of cultural knowledge entails coming to engage the world in newly instrumented ways, achieving these new ways depends on particular heuristics, such as pattern searching and cognitive offloading, that individuals draws on so as to empower their contextual grip on the world (Dreyfus & Dreyfus, 1999). That is, individuals do not see artifacts per se but rather tacitly *enframe* the world as *ad hoc* equipment for achieving their contextual goals (Heidegger, 1977). In turn, by virtue of learning to use this equipment, new views emerge for the learner in the form of utilization schemas inherent to disciplinary content (cf. Bartolini Bussi & Mariotti, 2008; Sfard, 2002; Vérillon & Rabardel, 1995). Moreover, in the activity of applying these appropriated artifacts, people may reconfigure their strategies dramatically in ways that experts recognize as pedagogically desirable (Abrahamson, Gutiérrez, Lee, Reinholz, & Trninic, 2011). This conceptualization of mediated discovery as effortful yet somewhat inadvertent is epitomized in Vygotsky's essay on the birth of inscription:

> Gestures, it has been correctly said, are writing in air, and written signs frequently are simply gestures that have been fixed.…children frequently switch to dramatization, depicting by gestures what they should show on the drawing; the pencil-marks are only a supplement to this gestural representation.…Children do not draw, they indicate, and the pencil merely fixes the indicatory gesture. (Vygotsky, 1930/1978, pp. 107-108)

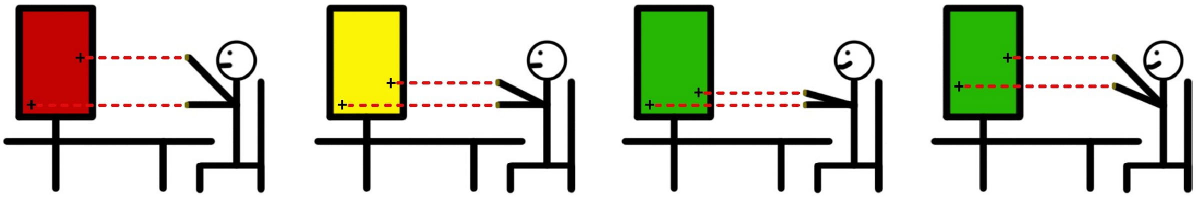
In this view, children's inscriptional productions ought to be conceptualized as initially no more than fortuitous signatures of projected gestures onto an available canvas—not as deliberate, stylized expressive depictions. Becoming literate thus consists of mastering the objectification of presymbolic notions in new semiotic systems embedded in a new media (cf. Radford, 2003; Saxe, 2004). We find this view of nascent literacy as transitional negotiations at the interface of semiotic systems particularly helpful in investigating technologically supported mathematics learning. Specifically, this paper scrutinizes empirical cases of students engaged in problem-solving activities who learn to appropriate virtual symbolic artifacts—a pair of crosshairs that mirror their hands on a computer display—as gestural extensions that amplify their inquiry. Building on sociocultural theory and its semiotic elaborations as well on cognitive and philosophical treatments of HCI, we ask, What interactive dimensions effectively characterize students' attempts to orient into virtual inquiry?

Data Source and Modes of Inquiry

The current study drew on data gathered in an ongoing design-based research study investigating the emergence of mathematical concepts from guided embodied interactional activities.

Our design conjecture, which built on the embodied/enactive approach (Lakoff & Núñez, 2000; Nemirovsky, 2003; Núñez, Edwards, & Matos, 1999), was that some mathematical concepts are difficult to learn because mundane life does not occasion opportunities to embody and rehearse the dynamic schemes that would form the requisite cognitive substrate for meaningfully appropriating the concepts' disciplinary analysis of situated phenomena. Specifically, we conjectured that students' canonically incorrect solutions for rational-number problems—"fixed difference" solutions (e.g., "2/3 = 4/5" - Behr, Harel, Post, & Lesh, 1993)—indicate students' lack of multimodal action images to ground proportion-related concepts (Pirie & Kieren, 1994).

Accordingly, we engineered an EI inquiry activity for students to discover, rehearse, and thus embody presymbolic dynamics pertaining to the mathematics of proportional transformation. At the center of our instructional design is the *Mathematical Imagery Trainer* (MIT; see Figure 1 and Figure 2; for detailed descriptions of the MIT's technical properties, see Howison, Trninic, Reinholz, & Abrahamson, 2011).



a. b. c. d.

Figure 1. The Mathematical Imagery Trainer (MIT) set at a 1:2 ratio, so that the right hand needs to be twice as high along the monitor than the left hand: (a) incorrect performance (red feedback); (b) almost correct performance (yellow feedback); (c) correct performance (green feedback); and (d) another correct performance.

a. b.

Figure 2. MIT in action: (a) a student's "incorrect" performance (raising both hands up away from a previous "green spot" at equal increments, i.e. fixed distance) turns the screen red; (b) "correct" performance (raising both hands up away from green at proportionate increments, i.e. different distances) keeps the screen green.

The MIT measures the height of the users' hands above the desk. When these heights (e.g., 10" & 20") match the unknown ratio set on the interviewer's console (e.g., 1:2), the screen is green. So if the user then raises her hands in front of this "What's-my-rule?" phenomenon by proportionate increments (e.g., to 15" & 30"), the screen will remain green but will otherwise turn red (e.g., to 15" & 25"; note that this pair of heights is 5" higher than both the left and right hands, respectively, in the 10" & 20" case). Study participants were initially tasked to "make the screen green." Once the participants successfully effected a green screen, they were tasked to "find another green"; then eventually tasked with keeping the screen green *while* moving their hands.

The initial condition for green was set as a 1:2 ratio, and no feedback other than the background color was given (see Figure 3b; this challenging condition was used only in the last six interviews). Once the students displayed a degree of competence with "finding green," the protocol introduced incremental layering of supplementary mathematical instruments onto this microworld. First, crosshairs were introduced that "mirrored" the location of participants' hands (see Figure 3c). Next, a grid was overlain on the display monitor to help students plan, execute, and interpret their manipulations and, so doing, begin to articulate quantitative verbal assertions (see Figure 3d). In time, the numerical labels "1, 2, 3,…" were overlain on the grid's vertical axis to help students construct further meanings by more readily recruiting arithmetic knowledge and skills and more efficiently distributing the problem-solving task (see Figure 3e).

Participants included 22 students from a private K–8 suburban school in the greater San Francisco area (33% on financial aid; 10% minority students). Care was taken to include students of both genders from low-, middle-, and high-achieving groups as ranked by their teachers. Students participated either individually or paired in a semi-structured clinical interview (duration of mean 70 min.; SD 20 min.).

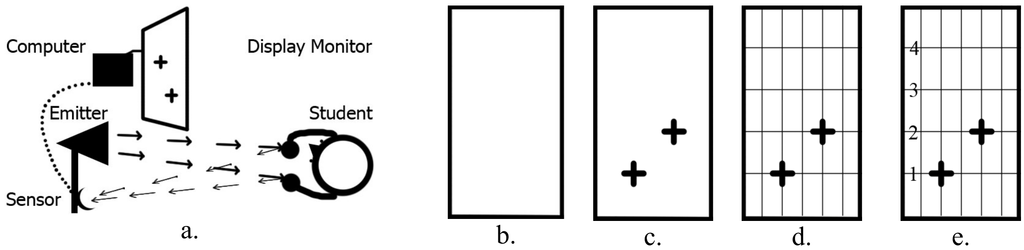


Figure 3. The Mathematical Imagery Trainer: (a) overview of the system featuring an earlier MIT version, in which students held tennis balls with reflective tape. Figures 3b – 3e are schematic representations of different display configurations, beginning with (b) a blank screen, and then featuring a set of symbolical objects that are incrementally overlain onto the display: (c) crosshairs; (d) a grid; and (e) numerals along the *y*-axis of the grid (in the actual design, the default range of the flexible grid and corresponding numerals was from 1 to 10).

Our investigation of the empirical data was conducted as collaborative, intensive microgenetic analyses (Schoenfeld, Smith, & Arcavi, 1991), and we applied general principles of grounded theory (Glaser & Strauss, 1967) to identify and articulate unfamiliar behavioral patterns in the entire data corpus. For this paper, we focus on two episodes, both of students navigating the early transition from a blank screen to a screen with interactive crosshairs (see Figures 3b&c, above; see our other cited publications for analyses of students' later behaviors). We view these episodes as paradigmatic cases of learning environments wherein crossing the physical-to-virtual divide is required for students to orient into a symbolically enhanced domain of scrutiny with vastly augmented inquiry and representational affordances. In particular, these data enable us to examine how students are able to render visible artifacts "invisibles" so that they can immerse themselves in thinking "through the computer."

Results: Two Data Episodes at the Gestural–Symbolic Interface

In this section, we present and discuss empirical data gathered around the introduction of the crosshairs. Whereas crosshairs per se are arguably *a*mathematical objects, their instrumentalization as virtual manipulation devices is a vital step towards progressive mathematization of the problem space, and, indeed, towards orienting into the problem space at all. As such, we view the introduction of crosshairs as instantiating a phenomenological breach of two dimensions, the embodied/physical/here ("I") and the inscriptional/virtual/projected ("it"). As we observe in the data, this breach and how the student and instructor attempt to ford it may render visible some of students' implicit framings of the situation. We now present two brief episodes: whereas the first showcases a user's capacity to learn a system by using it, the second illustrates a "lower boundary" of the same capacity.

Fording the Gestural–Symbolic Divide

Asa is a 5th-grade male student indicated by his teachers as low achieving. In this study, Asa was paired with Kaylen, another 5th grader likewise indicated. Sitting side by side in front of the technology, they each hold and manipulate one of the two tracker devices and attempt to co-produce a green screen. Prior to the layering of the crosshairs into their interaction space, the boys were observed to alternate their perceptual attention back and forth between the hand-held devices and the computer display. They looked at their devices so as to note their relative positions (e.g., to maintain or change the distance between them), and they looked at the screen so as to note the color feedback (whether it is green or red). Even when they lifted their hands to the level of their eyes, such that the devices were in the same line of vision as the screen, we still note the participants adjusting their gaze back and forth. JFG, a graduate-student researcher apprenticing through this study, facilitated this activity. The following transcript captures Asa's utterance *prior* to the introduction of the crosshairs.

Asa: <04:59> Oh I see. So I think what's going on is that they [referring to hand-held devices] have to be the same far… [RH thumb and index gestures a vertical interval] the same distance away from each other [RH, still holding the interval, gestures toward Kaylen's device].

Above, we note that Asa conjectured what we have called a "fixed distance" interaction conjecture, namely Asa believes that the distance between the hands should be invariant (see, below, Asa's "theory"). We also note Asa's use of the pronoun "it" in "…keep it the same distance," which suggests Asa has objectified the vertical interval between the two devices as the thing to be controlled. JFG now revealed the crosshairs on the screen.

JFG: <12:26> Can you see that?
A/K: Yeah. [Some technology recalibration ensues for a few seconds]
JFG: So let's see if these help us. [...] See if you can make the screen green.
Asa: Well... [gazing at screen] So let's try to find green again. There. Stop! Yeah. I think my theory's right. Oh wait, but we both have to be *moving*, 'cause when we stop...
JFG: You both have to be moving? What do you…What do you mean by that?
Asa: Seems so... Oh wait, no, that can't be possible. [to Kaylen] Stop! [screen is still green] Yeah it's not true. Go up.

Asa temporarily considered a new theory, by which the crosshairs must be in continuous motion. He then quickly refuted this theory in light of the empirical data they gathered.

JFG: <13:55> You think you can find another green [gestures higher on screen] kind of up here…?
Asa: My theory is obviously wrong.
JFG: What was your theory before?
Asa: My theory was that there was a specific [RH thumb/index again gestures vertical distance] height that they have to be from each other but [waves hand] that's wrong [shakes head].
JFG: So what do you… What do you think now?
Asa: Uhhm, I think that the height slowly increases. [Kaylen concurs]

In contrast to their alternating gazes back and forth between the hands and screen prior to overlaying the crosshairs on the screen, once the crosshairs appeared the children instantaneously gazed at them and thereafter never looked at their handheld devices while operating them. This transition from *here* to *there* led Asa to re-examine his previous fixed-distance conjecture; in turn, the increased precision afforded by this transition enabled Asa to gather data that he interpreted as rendering his conjecture "obviously wrong." Furthermore, the gestural–symbolic transition revealed one of Asa's implicit framings; namely, that devices/crosshairs "both have to be *moving*" so as to effect a green screen. At the point in the interview when the crosshairs were introduced, Asa had interacted with the artifact for approximately ten minutes with only modest progress along the dyad's inquiry into the mystery device. Then, over a very brief period of time *and without any expert modeling*, Asa was able not only to disprove two of his conjectures (fixed distances, moving), but also to produce a new conjecture, correctly noticing that the distance between the crosshairs should increase as they move up. In sum, Asa rapidly and seamlessly forded the gestural–symbolic divide and immediately availed successfully of the supplemental inquiry affordances in this virtually extended interactive problem space.

Stuck in the Gestural–Symbolic Divide

Next, we discuss the case of Boaz, a 5th-grade male student indicated by his teachers as "low achieving." The text below is a transcription from the early part of the interview, before the crosshairs had been revealed on the display. Boaz has been attempting to make the screen green. At the beginning of the transcript, he locates a "green pair" on the bottom of the screen. DR—another graduate-student research apprentice acting as interviewer—asks Boaz to explain his green-making strategy. In his response, Boaz will be referring to a "camera." This is not the video camera recording the session, but a sensor mounted on a tripod, part of the remote-action technology that picks up the infrared signal in the LED beam emitted by the handheld device the child operates. As we shall see, Boaz's idiosyncratic fixation on this device impeded his incorporation of the virtual objects and thus delimited the scope and efficacy of his inquiry into the mystery device.

DR: <14:02> Well, then, *how* did you make it green? [gestures toward computer display]

Boaz: *Oh!* By... uhhm…

DR: Are you figuring the green?

Boaz: By, uhhhm... [swerves body to the left, orienting torso away from the large monitor and toward the sensor; points toward sensor with both palms] By… So, I guess these two things right here [points to the devices that he placed on the desk] are showing the, uhmm [points to sensor] camera. Like [points back at devices] they're looking at the camera, and, I guess, those light things on it [refers to an obsolete LED array that is still mounted to the sensor, a relic of an earlier design cycle] each have a different color in each column. And so when they move it up or down or equal [mimes accordingly], each one, like, I'm, it's pinpointing….

DR: <15:23> So what would you do to make it green?

Boaz: I would... [operating in the lower range of the interaction space, just above the desk, Boaz lifts and lowers the two devices independently, monitoring for color on the large display; he stops when the screen turns green and holds his hands fixed at that position] …try pinpointing it by moving them up and down 'til I made green.

DR: Ok. So…[DR takes a mimed cue from DA] What if you tried to lift your hands up like this [lifts both hands *beside each other* into the upper ranges of the interaction space—a "neutral" gesture meant to mask actual locations] and keep it green? Do you know any way to do that?

Boaz: [Boaz imitates by lifting his hands up *beside each other* above his shoulder height, still staying in the interaction space. The screen turns red] Oh! It would go out of the motion [gestures toward DR's laptop]. It would go out of the camera [gestures toward sensor].

In the course of explaining to Boaz the boundaries of the physical interaction space, DR mimed contiguous hand motions, which would not effect a green screen. Boaz, however, does not share DR's framing, and so he cannot differentiate between DR's intended phenomenal element of vertical up-and-down motion and epiphenomenal or tactical performance element of contiguous hands. Boaz thus literally imitates the contiguous motions—he assumes he was supposed to do just so, so that when he effects a red screen rather than the anticipated green screen, he concludes the approach must be *technically* in error, i.e., that he is out of detectable range. Surely, thinks Boaz, the teacher showed him precisely how to make the screen green, so that the red screen could not have resulted from the teacher's error but from something else. This is incorrect, we note, as in fact the devices remained well within the interaction space: the screen turned red because DR's instructive gestures deviated from the desired 1:2 ratio. Thus Boaz inadvertently conflated the system's technical and mathematical aspects—a legitimate, rational inference, but one that impeded his performance for the remainder of the interview.

Later on in the interview, when the crosshairs were eventually revealed, Boaz remained physically and perceptually oriented toward the sensor. Albeit he began tracking the crosshairs in his peripheral vision, his alternating "I"/"it" utterances suggest he wavered between his hands and projection thereof, between a gesture and its inscription, and thus between the situation's technical and mathematical aspects. In contrast to Asa, Boaz remained "stuck" between the two media, greatly impeding his chances to discover the embedded mathematics.

Discussion

In opening this paper, we stated that is was about theory-meets-design. It is particularly about embodied-cognition theory meets design-based research. Although both the theoretical models of embodied cognition and the methodological paradigms of design-based research are on the rise in the mathematics education community, critical questions remain unanswered regarding the interaction of the two—about implementation of embodied-cognition theory in practice and the use of design to evaluate and produce embodiment theory. We hold that one promising approach to understanding how these theory and design relate involves examining how users learn to extend their inquiry into virtual realms through appropriating the affordances of new EI systems.

Having introduced our design, we turned our attention to two episodes drawn from a recent design-based research study that investigated the emergence of proportional reasoning from EI activity. The first episode showcased how the introduction of crosshairs on the screen enabled a student to learn how a system works by using it. The second episode, in contradistinction, presented the case of a student who remained stuck between what we deem technical and mathematical dimensions of the situation to the detriment of his inquiry and learning. Whereas this latter student was the only such case in our data, this exception affords greater insight into the rule. How might we make sense of Boaz's inability to learn how a system works by using it?

The introduction of the crosshairs marks a critical incipience in the hybridization of the embodied and inscriptional spaces (cf. McLuhan, 1964)—when users first realize that their gestures are tracked in real time, they experience the crosshairs as functional extensions of their manipulations (cf. Gangopadhyay & Kiverstein, 2009). We believe that, akin to children in Vygotsky's study who, whilst drawing, transition from gesture to inscription, the introduction of crosshairs presents EI users with an opportunity to transition to a new medium. Children's drawings both mark their gesture and serve to create images; crosshairs likewise bear *two* potential affordances: (1) interaction feedback for monitoring remote-interaction traction (i.e., "you are here"); and (2) strategic manipulation of the environment in accord with task demands (i.e., maintaining green). As designers, we rank these two maintenance goals respectively as: (1) peripheral (technical); and (2) central (e.g., mathematical). However, the designer's discerning perspective does not necessarily transfer to the child, who perceives the entire system as one new whole. Indeed, our data suggest that whereas Asa reconciled the two maintenance goals, Boaz did not. On the contrary, Boaz's actions indicate that, while utilizing the crosshairs' affordance to indicate remote-control traction, he misses out entirely on their pivotal *strategic* relevance in effecting green. This inability to differentiate between central and peripheral utility is not unlike that of children in Vygotsky's study, who "[w]hen asked to draw good weather…indicate the bottom of the page by making a horizontal motion of the hand, explaining, 'This is the earth,' and then, after a number of confused upward hatchwise motions, 'And this is good weather'" (Vygotsky, 1930/1978, p. 108). To these children, as to Boaz, the central activity is in the gesture, not the projected inscription.

Vygotsky's example of a child representing good weather suggests that Boaz's behavior, an exception in our particular study, may in fact be learners' modus operandi at the interface of physical and virtual media. More generally, it appears that users cannot learn how a system works by using it if they do not have some initial sense of what the system *is*—if they have not even begun to access its intended envelope of interactivity. And yet, systems such as the MIT interactive device afford as interactive potential not only designer-intended schemes, but also a host of unintended schemes, such as fixating on peripheral interaction (cf. Olive, 2000).

This, of course, is not meant to discourage EI design. Indeed, the remote-control technology giving rise to EI design bears an implicit yet profound affordance for instantiating progressive mathematization. In particular, it enables students to cross the embodied–inscriptional divide that is vital for making mathematical sense of immersive body-based extended functionalism. For example, crosshairs as virtual objects that mirror students' hand positions bear an ambiguous semiotic status: students can experience these objects either as functional extension of the body (tips of "invisible canes") or as data points expressing their embodied inquiry in a virtual reference field of measurement. Boaz is like a blindfolded person who has been offered a cane but has learned to use it only as a means of avoiding obstacles and not yet to "see" let alone engage the obstacles. In other words, whereas Asa incorporated crosshairs into his reasoning and, thus, reasons *through* them, to Boaz the crosshairs remain opaque, merely extensional location markers.

We are now able to provide some preliminary answers to the research questions guiding this work. First, what are the unique affordances of EI design for mathematical learning? By phenomenalizing proportion into an interactive mystery device, we rendered mathematics learning as akin to empirical scientific inquiry. That is, in order to make the screen green, students need to figure out empirically the principles governing this interactive phenomenon that comes to exist only by doing. Second, how can EI design aid us in developing and articulating theory? EI design demands that users hybridize the embodied and inscriptional, and the consequent interaction transition lays open to scrutiny hitherto tacit cognitive activity. Namely, this hybridity presents itself as a double-edged sword: on the one hand, it provides users with an opportunity to link their inscription to gesture; on the other hand, the divide between the embodied and the inscriptional may prove too great—it may bog a student in the doldrums of the intended system's peripheral features.

Yet our analysis offers insight toward answering another, third question: How might theory influence EI design? With respect to the issue of crossing the embodied–inscriptional divide, we advocate a markedly low-tech approach. Namely, we recommend looking back at the seminal work of Lev Vygotsky and his collaborators. They have already articulated the historical human achievement of fording the gestural–inscriptional media divide in the case of drawing on paper. This theoretical precedent, we submit, still holds for future studies into implementing equally successful media artifacts in the electronic age.

Whereas the empirical work explored here involved a relatively small group of children and treated brief interactions, the results suggest not only the methodological value of embodied-interaction design, but also future directions towards the development of effective computer environments for mathematics learning.

References

Abrahamson, D., Gutiérrez, J. F., Lee, R. G., Reinholz, D., & Trninic, D. (2011, April). *From tacit sensorimotor coupling to articulated mathematical reasoning in an embodied design for proportional reasoning.* Paper presented at the annual meeting of the American Educational Research Association (SIG Advanced Technologies for Learning). New Orleans, April 8 - 12.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617-645.

Bartolini Bussi, M. G., & Mariotti, M. A. (2008). Semiotic mediation in the mathematics classroom: Artefacts and signs after a Vygotskian perspective. In L. D. English, M. G. Bartolini Bussi, G. A. Jones, R. Lesh & D. Tirosh (Eds.), *Handbook of international research in mathematics education, 2nd revised edition* (pp. 720-749). Mahwah, NG: Lawrence Erlbaum Associates.

Behr, M. J., Harel, G., Post, T., & Lesh, R. (1993). Rational number, ratio, and proportion. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 296-333). NYC: Macmillan.

diSessa, A. A. (2008). A note from the editor. *Cognition and Instruction, 26*(4), 427-429.
Dreyfus, H. L., & Dreyfus, S. E. (1999). The challenge of Merleau-Ponty's phenomenology of embodiment for cognitive science. In G. Weiss & H. F. Haber (Eds.), *Perspectives on embodiment: the intersections of nature and culture* (pp. 103-120). New York: Routledge.
Gangopadhyay, N., & Kiverstein, J. (2009). Enactivism and the unity of perception and action. *Topoi*, 28(1), 63-73.
Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: strategies for qualitative research*. Chicago: Aldine Publishing Company.
Heidegger, M. (1977). *The question concerning technology and other essays* (W. Lovitt, Trans.). New York. (Original work 1954): Harper and Row.
Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): a framework for perception and action planning. *Behavioral and Brain Sciences, 24*, 849-878.
Hornecker, E., & Dünser, A. (2009). Of pages and paddles: children's expectations and mistaken interactions with physical-digital tools. *Interacting with Computers, 21*, 95-107.
Howison, M., Trninic, D., Reinholz, D., & Abrahamson, D. (2011). The Mathematical Imagery Trainer: from embodied interaction to conceptual learning. In G. Fitzpatrick & C. Gutwin (Eds.), *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems (SIGCHI 2011).*
Karmiloff-Smith, A. (1988). The child is a theoretician, not an inductivist. *Mind & Language, 3*(3), 183-195.
Lakoff, G., & Núñez, R. E. (2000). *Where mathematics comes from: how the embodied mind brings mathematics into being*. New York: Basic Books.
Leont'ev, A. N. (1981). The problem of activity in psychology. In J. V. Wertsch (Ed.), *The concept of activity in soviet psychology* (pp. 37-71). Armonk, NY: M. E. Sharpe.
Marshall, P., Cheng, P. C.-H., & Luckin, R. (2010). Tangibles in the balance: a discovery learning task with physical or graphical materials. In *Proceedings of the 4th ACM International Conference on Tangible, Embedded, and Embodied Interaction* (pp. 153-160). NYC: ACM.
McLuhan, M. (1964). *Understanding media: the extensions of man*. New York: The New American Library.
Nemirovsky, R. (2003). Three conjectures concerning the relationship between body activity and understanding mathematics. In R. Nemirovsky & M. Borba (Coordinators), Perceptuo-motor activity and imagination in mathematics learning (Research Forum). In N. A. Pateman, B. J. Dougherty & J. T. Zilliox (Eds.), *Proceedings of the Twenty Seventh Annual Meeting of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 105-109). Honolulu, Hawaii: Columbus, OH: Eric Clearinghouse for Science, Mathematics, and Environmental Education.
Núñez, R. E., Edwards, L. D., & Matos, J. F. (1999). Embodied cognition as grounding for situatedness and context in mathematics education. *Educational Studies in Mathematics, 39*, 45-65.
Olive, J. (2000). Computer tools for interactive mathematical activity in the elementary school. *International Journal of Computers for Mathematical Learning, 5*(3), 241-262.
Pirie, S., & Kieren, T. E. (1994). Growth in mathematical understanding: How can we characterize it and how can we represent it? *Educational Studies in Mathematics, 26*, 165-190.
Radford, L. (2003). Gestures, speech, and the sprouting of signs: A semiotic-cultural approach to students' types of generalization. *Mathematical Thinking and Learning, 5*(1), 37-70.
Raymaekers, C. (2009). Editorial on enactive interfaces. *Interacting with Computers, 21*, 1-2.
Salomon, G., Perkins, D. N., & Globerson, T. (1991). Partners in cognition: extending human intelligences with intelligent technologies. *Educational Researcher, 20*(3), 2-9.
Saxe, G. B. (2004). Practices of quantification from a sociocultural perspective. In K. A. Demetriou & A. Raftopoulos (Eds.), *Developmental change: theories, models, and measurement* (pp. 241-263). NY: Cambridge University Press.
Schoenfeld, A. H., Smith, J. P., & Arcavi, A. (1991). Learning: the microgenetic analysis of one student's evolving understanding of a complex subject matter domain. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 55-175). Hillsdale, NJ: Erlbaum.
Sfard, A. (2002). The interplay of intimations and implementations: Generating new discourse with new symbolic tools. *Journal of the Learning Sciences, 11*(2&3), 319-357.
Stetsenko, A. (2002). Commentary: Sociocultural activity as a unit of analysis: How Vygotsky and Piaget converge in empirical research on collaborative cognition. In D. J. Bearison & B. Dorval (Eds.), *Collaborative cognition: children negotiating ways of knowing* (pp. 123-135). Westport, CN: Ablex Publishing.
Vérillon, P., & Rabardel, P. (1995). Cognition and artifacts: a contribution to the study of thought in relation to instrumented activity. *European Journal of Psychology of Education, 10*(1), 77-101.
Vygotsky, L. S. (1930/1978). *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press.

How Are Students' Problems Being Solved? The Quality of Worked Solutions on a Popular Open, Online, Mathematics Help Forum

Carla van de Sande, Arizona State University, carla.vandesande@asu.edu

Abstract: Free, open, online, help forums connect students with volunteer helpers who provide assistance with specific problems from coursework. One of the most popular existing mathematics help forums, Cramster, is an advocate of Cognitive Load Theory (CLT), and promotes the provision of worked solutions. The purpose of this project was to ascertain the quality of the worked solutions that students are receiving, by analyzing 128 threads from the Algebra thru Pre-Calculus site archives from two perspectives. First, a rubric based on research-based design principles of worked examples was used to distinguish poor from well-structured solutions. Second, student ratings of solutions they received were examined and compared with the research-based ratings. The results show that Cramster is providing students with responses that contain steps, and sometimes goals, as opposed to final answers only. However, there is also much misinformation in the responses, and students do not show much discernment in their ratings.

Introduction

Students often seek help understanding course material and completing their homework assignments outside of school hours. Free, open, online, homework help forums are an unregulated Internet-based resource that allow students to anonymously ask for help on specific problems using threaded discussions. Responses come from volunteers who have the time, desire, and (perhaps) experience to help. Many such forums exist for a wide variety of school subjects, and are especially popular for challenging, homework-intensive subjects like mathematics. One the largest forums, Cramster, can be found on www.cramster.com. Cramster is a proponent of Cognitive Load Theory (CLT), with influential educational researcher and theorist, John Sweller, as a member of its advisory board. Cramster's modus operandi is to give students access to worked solutions, and the site logo (CrA$^+$mster.com™, *problem solved*) reflects the conviction that this is a good – if not the best – way to help students learn. At the same time, Cramster recognizes that students are responsible for using the worked solutions they receive in a productive manner: "The Cramster.com team is comprised of current and former students and we know that while copying answers on Cramster.com will get your homework finished, it will not improve your understanding of a subject."

The purpose of this paper is to explore the nature of the worked solutions that students are receiving on Cramster in the context of Algebra thru Pre-Calculus. Due to the unregulated nature of the site, the responses might well consist of final answers only. On the other extreme, there is the possibility that, in accordance with Cramster's philosophy, the responses look very much like worked examples commonly found in textbooks or other instructional materials. As part of an ongoing investigation on help forum participation (van de Sande & Leinhardt, 2007, 2008; van de Sande, 2010, in press), this paper views the responses students receive on Cramster through the lens of cognitive research on worked-out examples, and also takes into account how the students themselves view the quality of the help they are receiving. To help situate the reader, a brief description of Cramster is first provided.

Cramster

Cramster is a global online study community for high school and college school subjects that was launched in 2003 by Cramster, Inc., a private company headquartered in Pasadena, California. Cramster offers spontaneous online help, in the sense that *any member can contribute* to the community, *regardless of their qualifications*. In addition to the Q&A board (homework help), the site provides access to study materials such as worked solutions to exercises found in popular textbooks, topic notes, sample problems, and practice exams that were created by the administration, indexed from the web, or contributed by members. Although general membership is free, participants enjoy additional benefits and access (e.g., ability to ask more questions per day on the Q&A board) if they subscribe monthly ($9.95/month) or annually ($49.95/year). The Q&A board has four sections (Math, Science, Engineering, and Business) with subforums organized by subject area (such as Algebra thru Pre-Calculus). Members have access to user profiles that include self-volunteered information on school, major, and expertise. In each contribution, members are characterized by a self-selected username, an optional avatar, board level (from Rookie to Oracle), and "Karma." Karma points are acquired through participation in site activities (such as answering queries on the Q&A board, submitting solutions to textbook problems, challenging others' solutions, providing notes, quizzes, outlines, etc., and referring friends to become members), and are based on an intricate point system. In the Q&A board, Karma points depend on the rating awarded by the student who asked the question (Lifesaver, Helpful, Somewhat Helpful and Not Helpful) and the difficulty level of the board. For instance, answering a question on the Algebra thru Pre-Calculus forum posed by a non-paid/gold/platinum member could net 3/4/5 points, respectively, if the response receives a "Lifesaver" rating from the student who posed the request within 7 days. Similarly, Helpful and Somewhat Helpful ratings would net Karma points, just fewer. Karma points can be redeemed for gift certificates and merchandise, ranging in value from t-shirts to computers. On the part of students, rating responses is voluntary but does affect the student's "respect score" (the number of answers they have rated divided by the number of answers provided to them). This policy encourages students to rate the help received since students with low respect scores are presumably less likely to acquire (timely) help from forum helpers who are seeking to maximize their Karma points. Students can also leave comments, addressing a helper's response.

Cramster has recently (July, 2010) altered the look and feel of their site, together with many of their operative policies. The policies described above are the ones that were in effect at the time of the data collection. The current policies and their intentions were discussed in a phone interview with the Cramster press contact (Carleigh McKenna, personal communication, September 15, 2010). In the newer version of Cramster, instead of being able to post a specific number of questions per day depending on subscription type, students must earn or purchase Karma points in order to post questions. This policy discourages students from opening multiple accounts with different user names, and, at the same time, encourages students to actively contribute to the community (and thereby earn Karma points so that they can have their needs met). Students also now specify a time frame in which they need to receive a reply ("fast" for 12-24 hours, "faster" for 4-12 hours, and "fastest" for 0-4 hours), with each of these designated response speeds "costing" students progressively more Karma points. In addition to introducing variable point transactions for speed, the Karma point system was recalibrated, so that more Karma points (roughly 10 times as many) are at stake for a given service. Finally, the rating category "Somewhat Helpful" has been altered to "Needs More Work." The hope is that this change in semantics will encourage communication between students and helpers, so that helpers will make clarifications when needed and students can get additional help if the problem has not been resolved for them.

Theoretical Perspectives

Although help seeking once carried with it a stigma of laziness or incompetence, many theorists now recognize this activity as a potentially legitimate part of the learning process (Karabenick, 1998; Nelson-Le Gall, 1985; Newman, 1994), and one that underscores the Vygotskian (1978) notion of learning as inherently social. Whether or not help seeking has strategic value for learning depends on several attributes of the situation (Karabenick & Newman, 2006), including the performance goals of the participants. If the goals are mastery oriented, then the help-seeking encounter is instrumental and can provide a means of acquiring new knowledge and skills. On the other hand, if the goals are dependency oriented, then the help-seeking encounter is not a legitimate learning activity and instead merely serves as a means of acquiring information or solutions (Nelson-Le Gall, 1981).

It is within this theoretical perspective of help seeking that theories of learning come into play. Different learning theories offer us different definitions of what constitutes effective help and how best to promote mastery-oriented help seeking, particularly in the context of problem solving. Cognitive load theory (CLT), in particular, offers several extensively researched principles for the design of instructional materials (Mayer, 2005). Taking into account the limited capacity of working memory, CLT emphasizes the need for instructional design to minimize extraneous cognitive processing or load and to maximize germane load (Sweller, 1989; Sweller, van Merriënboer, & Paas, 1998). The argument is that certain practices contribute to extraneous cognitive load, making it more difficult than necessary for students to learn and retain the material, whereas other strategies lower extraneous cognitive load and allow all working memory capacity to be devoted to schema construction.

The use of worked examples is one practice that is consistent with reducing extraneous load, and that, therefore, CLT strongly advocates. The provision of worked examples as a study tool has been shown to facilitate near transfer, to shorten acquisition and performance time, and to make learning seem less effortful than engaging in problem solving or exploration (Paas & Van Merriënboer, 1993; Sweller & Cooper, 1985; Tuovinen & Sweller, 1999; van Merriënboer, Schuurman, De Croock, & Paas, 2002). These benefits are especially pronounced for students who are unfamiliar with the material that is being learned (Kalyuga, Chandler, Tuovinen, & Sweller, 2001).

However, worked examples are not all equally effective. The structure of worked examples matters, and "may substantially compromise the benefits derived from studying them" (Mwangi & Sweller, 1998, p. 174). Higher quality examples provide cues to students regarding the integration of various components (Atkinson, Derry, Renkl, & Wortham, 2000). In particular, examples in which goals and subgoals are highlighted, using labels or visual chunking, are more easily learned from than examples presented as a string of steps (Catrambone, 1995, 1998).

Methods

The Corpus

All of the threads from the archives of the Algebra thru Pre-Calculus Q&A board were collected for the month of September, 2009. This netted 1,246 threaded discussions. Next, in order to have a tractable number of threads to analyze regarding response quality, 10 threads from each day were randomly selected, resulting in 300 threads. Of these, 235 were deemed topic appropriate for Algebra thru Pre-Calculus. The Algebra thru Pre-Calculus threads were next coded for the type of response sought according to problem type: Solution (worked solution to given exercise), Explanation (explanation or description of concept), Clarification (clarification of solution retrieved from other source), Verification (assessment of accuracy for solution), and Information (formula). Because this project focuses on the quality of worked examples, only the 211 threads that involved a student seeking a *solution* were selected. There were 32 threads in which, against Cramster policy, a student posed more than one exercise, and these were discarded. Finally, only threads in which the problem statement, as presented by the student in the initial post, was well-defined (e.g., complete and unambiguous) were included. After this culling, the data set amounted to a total of 128 student requests for solutions and 194 helper responses.

The 128 student requests came from 82-94 different students, and the 194 responses came from 109-115 different helpers (since 13 students and 7 helpers used the shared user name “Anonymous”). All user names referred to in this paper have been altered to protect the (online) identity of participants, but an effort was made to use pseudonyms that still convey the user’s persona.

The Analyses

All responses received to the 128 student requests for solutions were examined for quality from two perspectives: cognitive research-based principles for worked examples, and student perception.

Research-based Quality

To assess the quality of the help received on Cramster in light of theory-based design principles for worked examples, each helper response was assigned a rating from 0 to 5. Table 1 contains a description of some of the features that differentiated helper solutions. A 0 was assigned to helper responses that were either absent (0NR), were not a solution to the problem statement (0NS), or were mathematically inaccurate (0WR); a 5 was assigned to responses that almost had a “textbook quality” feel to them, clearly and concisely laying out the solution using steps, associated (sub)goals, labels, and, perhaps adding warnings or checks. Inter-rater agreement was 89% and all differences were resolved following discussion.

Table 1: Research-based quality rubric for solutions posted on Cramster.

Rating		Features
0	0NR	No response
	0NS	Response not representing solution to problem posed (e.g., solution to alternative problem statement)
	0WR	Mathematically inaccurate (wrong) response
1		Mathematically accurate response, but one that consists of a final answer only
2		Accurate response that provides a sketch of the solution, or an incomplete worked solution
3		Accurate response that closely resembles an instructor solution manual, containing some steps, but few, if any, (sub)goals or explanations
4		Accurate response that contains many steps and associated (sub)goals, but that is somewhat hard to follow (such as using inconsistent notation), or that contains extraneous steps (such as unnecessary conversions)
5		Accurate response that contains steps, associated (sub)goals, labels, and, perhaps, warnings or checks, with a clear and concise presentation style and format

Student-based Quality

Within 7 days of a response, students may rate the help on Cramster, as mentioned above. These ratings allow students to give the helpers feedback and provide an indication of how students view the quality of the responses. Because of the reformatting of the site, the original student ratings in the archives (whence the data set came) had been translated to the new rating system. There were 5 categories reflecting student pronouncement of quality, or lack thereof: Lifesaver (L), Helpful (H), Needs Work (NW) (old Somewhat Helpful), (NH) Not Helpful, and No Rating (NR).

Results

Research-based Quality

Figure 1 contains the percentage of responses according to their quality as worked examples using the research-based rubric. The quasi-bimodal distribution of the results tells the two sides of the story on the quality of the responses students are receiving on Cramster.

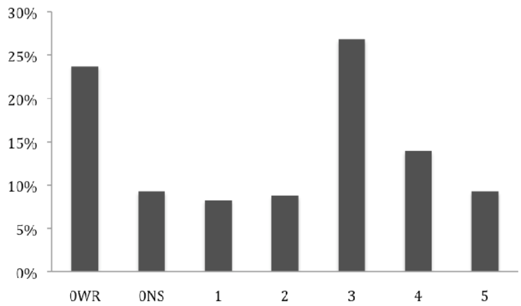


Figure 1. Percentage of Responses According to Quality as Worked Example.

The first part of the story is that students *are* receiving responses on Cramster, that the site is not just providing students with final answers, and, that many solutions posted are at least the quality of an instructor solution manual. Only 24 of the total 128 student requests in the data set, went unanswered (0NR). Although this is a relatively high nonresponse rate compared with similarly structured mathematics help forums (van de Sande & Leinhardt, 2007), it is still remarkable since Cramster helpers do not necessarily receive any compensation (Karma points) or commendation (forum status) for their efforts. The average number of responses to requests that were answered was approximately 1.9 (194 responses for 104 requests).

Furthermore, of the student queries that received responses, only approximately 7% of the responses consisted of a final answer only (1), resembling what one typically finds in a "back of the book" answer section. These responses are of absolutely no use as worked examples from which to learn, and violate Cramster's posted anti-cheating policy at that time: "Copying solutions or *posting unexplained final answers* on the Q&A Board promotes completion without comprehension, and that's something we don't support on this site" [italics added]. This low percentage is definite cause for comfort since the site is not regulated, standing outside the purview of formal educational institutions.

Finally, 44% of the solutions students are receiving on Cramster are the quality of those found in instructor solution manuals (3), or higher (4-5). These solutions, at the least, contain steps, and may also contain associated goals and explanations. Students could presumably use these solutions as worked examples to learn the material, much in the spirit of how they use examples that are provided in a textbook or other instruction materials. Figure 2 shows a high quality worked solution that was received in response to the problem statement: A 10-m ladder is leaning against a building. The bottom of the ladder is 5-m from the building. How high is the top of the ladder? The helper, Gondal, includes a labeled diagram, a goal ("use the pythaorean theorum"[sic]), an explanation ("right triangle"), and the solution broken down into steps. At the same time, this solution is not perfect; it contains several spelling errors, and, more importantly, does not explain why the negative solution value to the quadratic was not considered (e.g, because h is a length and therefore must be nonnegative). Despite these flaws, the student, What08, appreciated the quality of the response, giving it a Lifesaver rating and commenting "thanks."

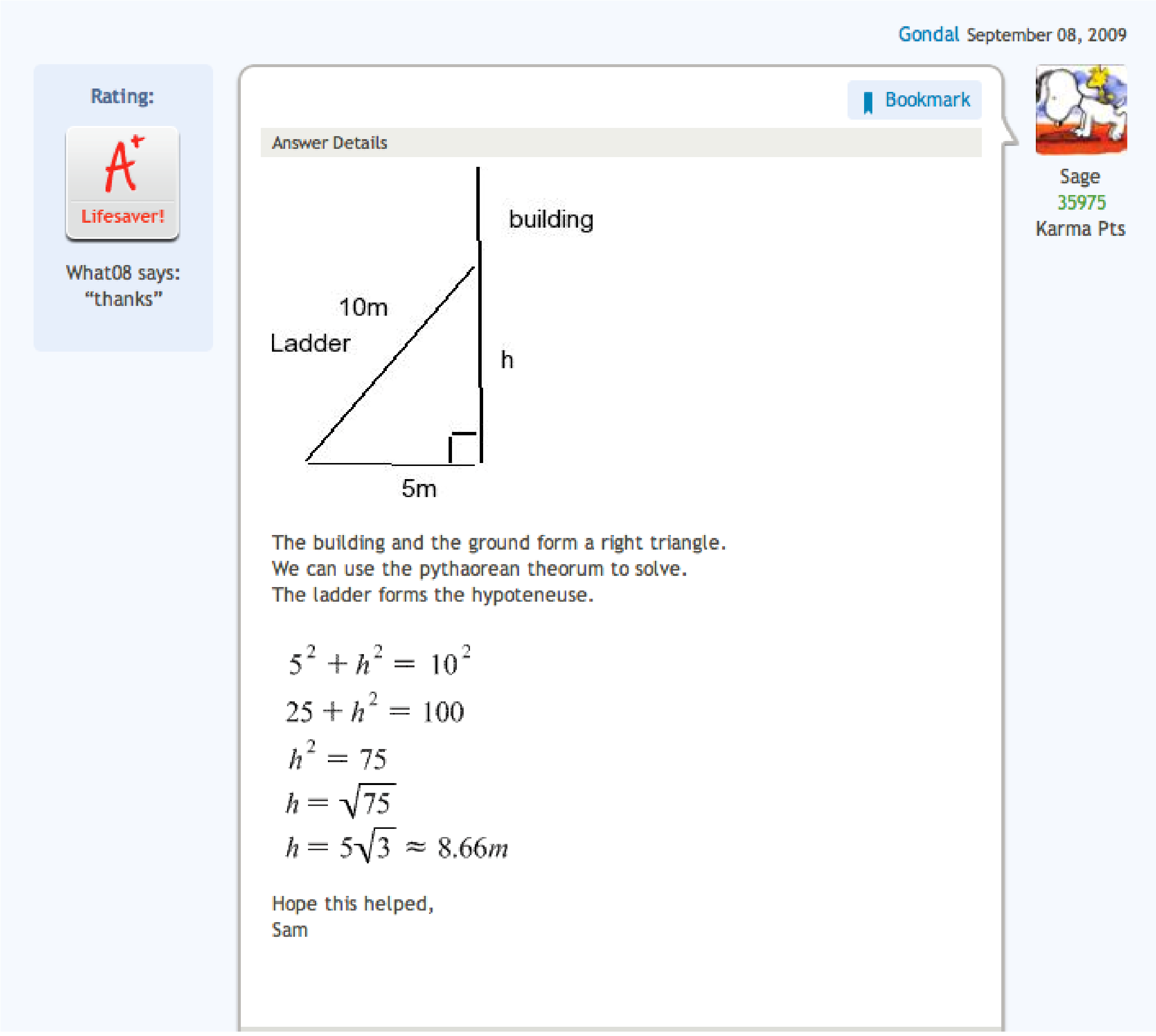


Figure 2. High Quality Worked Example Containing Steps, Explanations, and a Diagram.

The second part of the story, and a part that is cause for some concern, is that students are receiving incorrect information on Cramster, and that many of the responses they receive are not usable as worked examples. Indeed, 24% of the responses in the sample contained mathematically inaccurate statements (0WR). Here, I note that accuracy was coded very conservatively, and captured subtle errors that concerned mathematical aspects of the argument (such as the failure to use parentheses consistently), in addition to more flagrant transgressions (such as incorrectly factoring an expression). However, the justification for this decision was that the students who receive these responses cannot be assumed to have the mathematical experience necessary to detect errors (especially subtle ones) in an argument, and may be using these responses as a model.

Accuracy is the minimal criteria for being useful as a worked example. A number of responses also fell short of providing students with the semblance of a worked example. Altogether 41% of the responses received did not meet the criteria for being a passable worked example, either because they contained incorrect information (0WR), consisted only of the final answer (1), or, were in some way, incomplete (2).

Student-based Quality

Another metric for quality comes from the ratings given by student Cramster users to the responses they receive. Table 2 shows the number of responses (containing solutions) by student ratings (Not Helpful, Needs Work, Helpful, Lifesaver, No Rating) and the research-based quality rubric (0WR, 1-5).

As the last row in Table 2 shows, students did generally rate helper responses. Although they may have done so in order to maintain a high "respect score" (as discussed above), this score is completely independent of the actual student rating (a "Not Helpful" affects the student's respect score the same as a "Lifesaver" rating). Looking at Table 2, we see that, when they give them, students appear generous in their ratings. Of these responses that received ratings (n=163), the vast majority (n=156) received moderate (Helpful) to high (Lifesaver) marks, and no responses received the lowest rating, Not Helpful.

Table 2: Student ratings and research-based ratings for the quality of responses containing solutions.

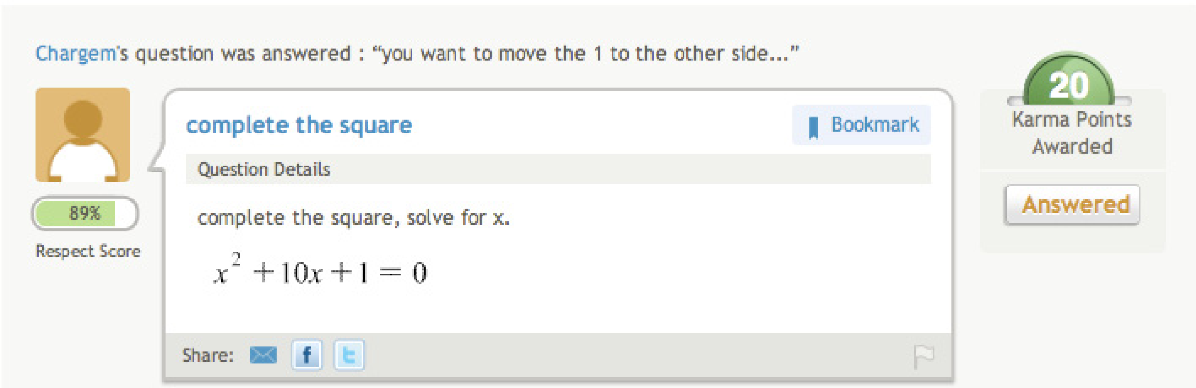
	Not Helpful	Needs Work	Helpful	Lifesaver	No Rating	Total
0WR	0	3	15	22	6	**46**
1	0	2	7	6	1	**16**
2	0	1	10	4	2	**17**
3	0	1	27	22	2	**52**
4	0	0	13	12	2	**27**
5	0	0	7	11	0	**18**
Total	**0**	**7**	**79**	**77**	**13**	**176**

Of course, the question that begs to be asked here is whether the marks are deserved, according to some independent and informed criteria, such as our quality rubric. The answer is that it appears that students are not very critical of "bad" help, and that they also appreciate "good" help. Of the solutions that were mathematically inaccurate (0WR), less than 7% (3 of 46) received poor marks. This trend also held for solutions that consisted of final answers only (1), for which only 13% (2 of 16) received a rating of "Needs Work." Solutions that were "good" worked examples (4-5) were recognized as such, receiving no poor marks (0 of 45) and equivalent Helpful (n=20) and Lifesaver ratings (n=23). The upshot is that students give high marks to inaccurate responses (55% received a Lifesaver rating, compared to 45% of the accurate responses). Also, students tend to be more prone to give higher ratings as the quality of the help increases.

An Example

Figure 3 shows an exchange in which an incorrect response received the highest rating possible and a correct response received a lower, although still good, rating. It is not clear what order the two responses were received in, since timing information, except for the date, is no longer published in the threads. The student, Chargem, gave the problem statement, which included a specific method, namely completing the square, and received a response from nzatisw58 that contained several quite serious mathematical errors. Indeed, the worked solution makes only one true statement (x^2+10x+1 = 0 can be rewritten as x^2+10x = -1). Chargem responded enthusiastically to this response, though, and, beyond giving it a "Lifesaver" rating, took the time to add an enthusiastic and appreciative comment for the "detailed explanation." Another response, from Firth, provided the correct solution (albeit with only one goal describing the first step for completing the square), but only received a "Helpful" rating without comment. Note, however, that neither helper alluded to a conceptual rationale for performing the steps that were included (e.g., a diagram illustrating the meaning of the terms).

Question:



Answers:

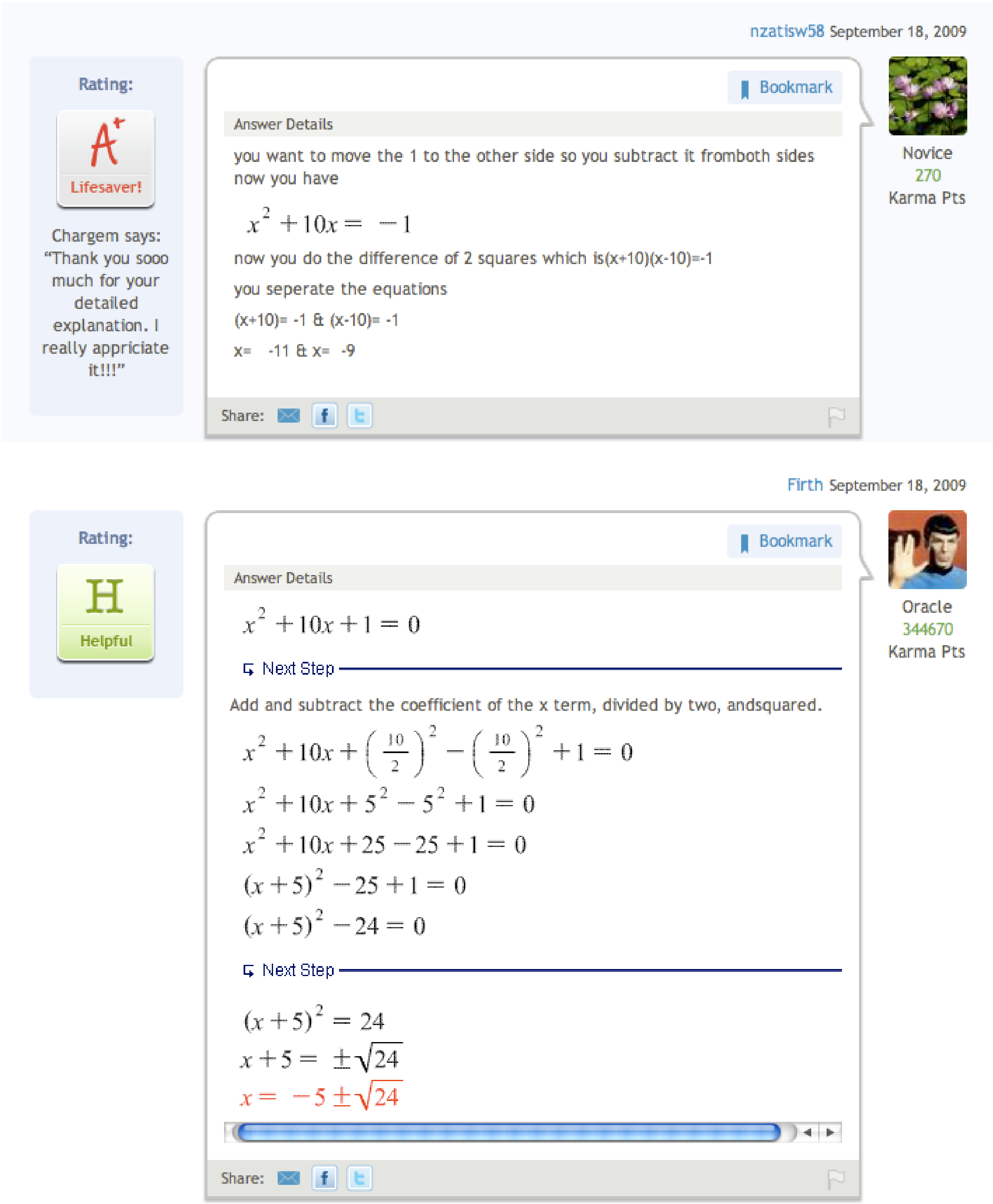


Figure 3. Question and Two Helper Responses: Incorrect one gets "Lifesaver." Correct response gets "Helpful."

Discussion

Thousands of students bring their problems to Cramster, presumably in the hopes of receiving a worked solution in good time from more experienced others. We see evidence from this project that students are generally receiving worked solutions as opposed to final answers, and that many of these worked solutions provide at least steps to follow. This is good news. On the other hand, many of the solutions also contain errors, some more serious than others. Such worked solutions could have the opposite effect than intended by giving unsuspecting and perhaps mathematically naïve students an incorrect example to follow as their guide. Based on their own ratings, it appears that students do not discern these errors and simply appreciate the fact that they have received assistance.

Of course, students may be communicating with helpers off-line (e.g., through private messaging) for clarification, and an observational methodology cannot capture this. Also, we cannot speak to the (ab)use of the responses as worked examples, regardless of their quality. Are students using the solutions they receive as problem solving models, or are they passively copying them? In the former case, students would be experiencing the benefits of learning from worked-out examples, whereas, in the case of copying, students are cheating and no learning is taking place. Finally, we have no evidence that providing worked examples in an online help forum environment trumps alternative pedagogical practices (such as guided discovery) favored by other learning paradigms. In sum, this project addressed *how* student's problems are being solved – the question of whether students' problems *are* really being solved still awaits an answer.

References

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research, *Review of Educational Research, 70,* 181-214.

Catrambone, R. (1995). Aiding subgoal learning: Effects on transfer. *Journal of Educational Psychology, 87,* 5-17.

Catrambone, R. (1998). The subgoal learning model: Creating better examples so that students can solve novel problems. *Journal of Experimental Psychology: General, 127,* 355-376.

Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Applied Educational Psychology, 93*, 579-588.

Karabenick, S. A. (1998). Help seeking as a strategic resource. In S. A. Karabenick (Ed.), *Strategic help seeking: Implications for learning and teaching* (pp. 1-11). Mahwah, NJ, Erlbaum.

Karabenick, S. A., & Newman, R. S. (Eds.). (2006). Help seeking in academic settings: Goals, groups, and contexts. Mahwah, NJ: Erlbaum.

Mayer, R. E. (Ed). (2005). *Cambridge Handbook of Multimedia Learning.* New York: Cambridge University Press.

Mwangi, W., & Sweller, J. (1998). Learning to solve compare word problems: The effect of example format and generating self-explanations. *Cognition and Instruction, 16,* 173-199.

Nelson-Le Gall, S. (1981). Help-seeking: An understudied problem-solving skill in children. *Developmental Review*, *1*, 224-246.

Nelson-Le Gall, S. (1985). Help-seeking behavior in learning. *Review of Research in Education*, *12*, 55-90.

Newman, R. S. (1994). Adaptive help seeking: A strategy of self-regulated learning. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulation of learning and performance: Issues and educational applications* (pp. 283-301). Hillsdale, NJ: Erlbaum.

Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1993). The efficiency of instructional conditions: an approach to combine mental effort and performance measures. *Human Factors, 35*, 737-743.

Sweller, J. (1989). Cognitive technology: some procedures for facilitating learning and problem solving in mathematics and science. *Journal of Educational Psychology, 81*, 457-466.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as substitute for problem solving in learning algebra. *Cognition and Instruction, 2*, 59-89.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design, *Educational Psychology Review, 10(3)*, 251-296.

Tuovinen, J. E., & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology, 91*, 334-341.

van de Sande, C. (2010, June). *Free, open, online, mathematics help forums: The good, the bad, and the ugly*. In K. Gomez, L. Lyons, & J. Radinsky (Eds.), Learning in the Disciplines: Proceedings of the 9th International Conference of the Learning Sciences 2010 (Vol. I, pp. 643-650). Chicago, IL: ISLS.

van de Sande, C. (in press). A Description and Characterization of Student Activity in an Open, Online, Mathematics Help Forum, *Educational Studies in Mathematics.*

van de Sande, C., & Leinhardt, G. (2007). Online tutoring in the Calculus: Beyond the limit of the limit. *Éducation et Didactique, 1(2),* 115-154.

van de Sande, C. & Leinhardt, G. (2008, June). *The Good Samaritan effect: A lens for understanding patterns of participation*. In P. A. Kirschner, F. Prins, V. Jonker & G. Kanselaar (Eds.), Proceedings of the 8th International Conference for the Learning Sciences 2008 (Vol. II, pp. 240-247). The Netherlands:ISLS.

van Merriënboer, J. J. G., Schuurman, J. G., De Croock, M. B. M., & Paas, F. G. W. C. (2002). Redirecting learners' attention during training: effects on cognitive load, transfer test performance and efficiency. *Learning and Instruction, 12*, 11-37.

Vygotsky, L. S., (1978). Mind in society: The development of higher psychological processes. Cambridge, Massachusetts, Harvard University Press.

Acknowledgments

This work was supported by a grant from the Spencer Foundation (#201000107).

Fostering Representational Competence through Argumentation with Multi-Representational Displays

Mike Stieff, Department of Chemistry, Learning Sciences Research Institute, University of Illinois, Chicago, mstieff@uic.edu

Abstract: The present study examines how students collaboratively reason about scientific phenomena by coordinating multiple representations presented in a single display in the Connected Chemistry modeling environment. A case study of two secondary students working together to solve a chemical equilibrium problem illustrates how students privilege one representation, yet problem solve by collaboratively integrating information from multiple representations. Working with the environment, individual students make a rapid initial judgment about which displayed representation is most useful for a particular task and ignore information presented in other representations unless their peers prompt them to use that information. Analyzed individually, each student would appear to lack representational competence, yet together the dyad demonstrates sophisticated coordination of representations. Using these observations, the study posits that an individual student's difficulties with coordinating representations can be mitigated by embedding multi-representational displays in collaborative activities that foster argumentation about representations.

Introduction

In many science domains, particularly chemistry, teaching and learning makes extensive use of external representations. In the classroom, students must often interpret and manipulate multiple representations of a single phenomenon to apprehend elementary concepts. For example, any given chemical reaction can be represented with an energy coordinate diagram, a molecular orbital diagram or a reaction mechanism that together describe the reaction. Historically, chemists have developed many representations to embed vast amounts of information in small diagrams and to make problem solving more efficient (Hoffmann & Laszlo, 1991). Indeed, the most successful problem solving strategies used by expert chemists employ simultaneous, coordinated use of several representations. Although unproblematic for experts, the wide variety of external representations and their coordinated use has been a reported learning barrier for chemistry students for decades (Johnstone, 1993). To address students' learning difficulties with multiple representations in chemistry, new computer-based curricula attempt to enhance students' interpretation of chemical representations through simultaneous displays of multiple representations (Kozma, Russell, Jones, Marx, & Davis, 1996; Stieff & Wilensky, 2003; Wu, Krajcik, & Soloway, 2001). Although each learning environment is unique, all share a common goal of making explicit the information embedded in multiple representations concurrently with a visual display of submicroscopic views of molecular interactions. Thus, each provides "multiple, linked representations" to help students coordinate among representations (Kozma, et al., 1996, p. 41).

Although such environments are based on the assumption that the simultaneous display of representations fosters learning, students have reported the same difficulties understanding and interpreting representations in these environments that they report when using canonical static representations. For example, in a study of the 4M:Chem learning environment Kozma and Russell (1997) reported that experts were able to use the displays intuitively and efficiently, but students were unable to choose among or coordinate the given representations to problem solve. Similarly, Wu et al. (2001) reported that students using the eChem modeling environment were given to comparing the surface features of displayed representations instead of comparing the embedded conceptual information. Thus, in the early stages of chemistry learning, students appear to lack content knowledge that allows them to make inferences from the wide array of chemical representations available. Clearly, the capabilities of novel technologies to display coordinated multiple representations is impressive, yet the historical barriers to working with multiple representations persist in this medium.

A close examination of student behavior when working with multiple representations offers some insight into the persistence of learning barriers in novel curricular environments. In a study of physics students, diSessa et al. (1991) posited that students maintain many ideas about the role of representations in science learning, which they termed "meta-representational competence". Namely, learning is heavily influenced by a student's "expertise in inventing, evaluating and refining a variety of representational forms" (diSessa et al, 1991, p. 86). Although this expertise is of course influenced by domain content knowledge, diSessa and colleagues have suggested that meta-representational competence is to some degree domain independent. Thus, students may be more or less given to making correct inferences with given representations or generating their own useful representations for problem solving. As such, simultaneous displays of multiple representations may present challenges for students less competent in evaluating and manipulating scientific representations.

Despite the reported challenges regarding (meta)-representational competence, recent work has revealed that students are quite capable of coordinating across multiple representations in computer displays to problem solving effectively. Using eye tracking techniques and verbal protocols of chemistry student problem solving with multi-representational displays, Stieff, Hegarty, and Deslongchamps (2011) illustrated that even with minimal content knowledge chemistry students are able to select the most effective representation to problem solve. Equally important, their results revealed that in cases where students do not verbally refer to using multiple representations, their eye movement behavior suggests they attend to information presented in multiple representations. The extent to which the students in their study made use of information in multiple representations remains unclear, however, as 80% of the participants relied primarily on the information presented in only one representation to justify their claims about chemical phenomena.

While work such as this demonstrates that students are capable of coordinating across representations in multi-representational displays, it fails to account for students' use of multiple representations in collaborative settings. Outside of chemistry, researchers have shown that multi- representational environments can be highly effective in supporting collaborative learning among students (Stahl, Koschmann, & Suthers, 2006; White & Pea, in press). In particular, such work has shown that designed environments that allow students to construct individual representations for collaborative use or designs that assign individual representations to group members and require group coordination often result in highly effective learning and problem solving. Together, these studies suggest that the challenges using multi-representational displays in the chemistry classroom can be overcome with carefully designed learning environments.

In this spirit, the present study analyzes representational competence displayed by a dyad using a multi-representational display in the chemistry classroom. Using a case study approach, I attempt to merge prior work in chemistry that centers on individual's working with multiple representations with work from other disciplines that focuses on the coordination of personal representations across groups. Specifically, I aim to explore how inquiry-based activities can encourage students to coordinate among representations using one shared multi-representational display. The results illustrate how individual students make a rapid initial judgment about which representation is most useful for making scientific claims and ignore information presented in other representations unless prompted by peers to use that information. Using these observations, the study posits that an individual student's difficulties with coordinating representations can be mitigated by embedding multi-representational displays in learning environments that include collaborative activities that foster argumentation.

Present Study

To examine representational competence with multi-representational displays, the present study analyzed six pairs of students completing one activity using the Connected Chemistry learning environment (Stieff, 2005). The activity discussed below included simulations programmed in the Netlogo modeling language (Wilensky, 1999). Using the simulation in the classroom, the dyads completed guided inquiry activities in which they observed molecular interactions to learn how macro-level concepts and relationships emerge from submicroscopic interactions. Here, I present observations from one dyad learning about chemical equilibrium.

Context & Participants

The case study presented here was constructed from field observations of one implementation of a Connected Chemistry Unit in Shadylane High School. Shadylane High is located in a middle-class urban community serving a primarily Asian and Latino student body and was ranked in the 7th decile statewide at the time of the study. Mr. Drake, who had worked in a chemistry-related industry for 20 years prior to teaching, had taught chemistry for 13 years at the time of this study. He noted that he rarely used educational technology resources in his classroom. Mr. Drake described the students in the case study, Jan and Sheila, as two "top achieving" and "highly motivated" students. Individual interviews with Jan and Sheila after the classroom observations reported here indicated that both students self-identified as college-bound and highly interested in science. Although neither student reported using computer activities frequently in the classroom, each had previously participated in a Connected Chemistry implementation several months prior to the present study. Thus, they stated they were familiar with the environment, and each had the sense that the computers were inherently helpful for learning chemistry. As Sheila said, the simulations "let you see what is going on".

During the three-day observation of Jan and Sheila working together on one laptop, the class completed a series of worksheets included in the Chemical Equilibrium Unit. The worksheets required the students to observe simulations of four different chemical reactions and answer questions about different aspects of each reaction. A primary learning objective of the unit aimed to support students' ability to distinguish *reversible* reactions from *irreversible* reactions. Briefly, *reversible* reactions are reactions that involve the simultaneous conversion of reactants into products and products back into reactants; *irreversible* reactions are reactions in which the reactants convert into products, but products do not convert back into reactants. To that end, each worksheet asked the students to observe both macroscopic plots and submicroscopic molecular interactions and make claims about which two reactions are reversible and justify their claim with evidence from the display.

The Chemical Equilibrium Simulation

Like all Connected Chemistry simulations, the Chemical Equilibrium simulation (see Figure 1) is a multi-representational display that includes three distinct representations that students are asked to coordinate. First, the simulation contains a *graphics window* that displays a simulation of submicroscopic chemical interactions using space-filling chemical representations. Second, the simulation includes a *plotting window* that displays macro-level variables, such as concentration, using a graphical representation. Third, the simulation contains symbolic representations of the chemical reactions displayed in the graphics window.

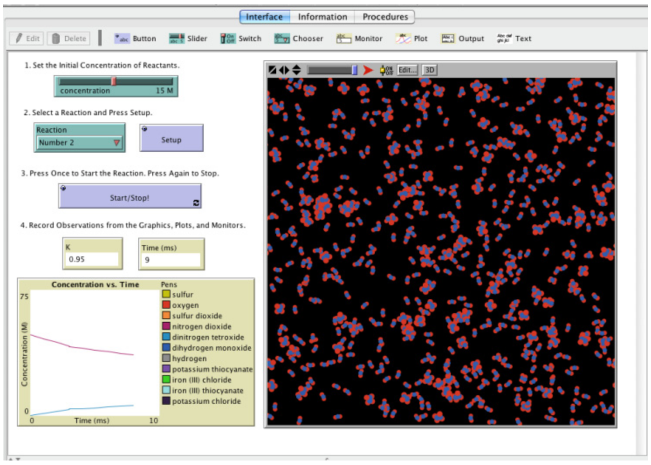


Figure 1. The Chemical Equilibrium #1 simulation includes a large graphics window that illustrates dynamic molecule interactions and a plotting window that illustrates macroscopic changes in concentration resultant from behavior in the graphics window. Symbolic representations are present in the legend and on the worksheets.

Coordinating Representations through Argumentation

Here, I discuss how one dyad decided which reactions were reversible while working on the Connected Chemistry Chemical Equilibrium Unit during the first day of the 3-day unit. The exchanges between Jan and Sheila reveal that each student privileges the information embedded in a unique representation on the display, but together they coordinate across representations by arguing about the information in each representation. I propose that the constraints of the activity to provide multiple pieces of evidence from the simulation to support a claim fostered disagreement between Jan and Sheila about which representation has the "best" evidence. In turn, this disagreement leads Jan and Sheila to coordinate their representations to justify their answers.

Sheila and Jan begin the first activity immediately after watching the 'Reaction 1' demonstration from Mr. Drake. Previously, Mr. Drake had described Sheila and Jan as two of his most motivated students, which was evident in their initial approach to the activity. The girls slowly read over the instructions independently and then together begin to manipulate the simulation for 'Reaction 2'. They note the activity asks them to record observations both from the plotting window as well as the graphics window to decide which two of the four reactions are reversible. From their first interaction with the simulation, Sheila appears to struggle interpreting the interface and the relationship between the graphics window, reporting monitors and the plot. In contrast, Jan rapidly reads information off the display. With Jan's support, Sheila learns to locate important information in the display, and in turn, teaches Jan how to find hidden information by interacting with the display (1).

Sheila: (*Reading from the worksheet.*) Reactants, what do we got? What is this? (*Sheila points to the graphics window.*) (...) I can't even tell what this is. I think it's oxygen?
Jan: I think it's N-O-2. (*Jan looks to where Sheila is pointing, and then looks at the molecular representation key on the worksheet. She leans forward and looks closely at the screen*). Yes, it's N-O-2.
Sheila: Are you sure?
Jan: Yes.
Sheila: OK, so nitrogen is/
Jan: That's the reactant that we started with.
Sheila: I guess so. So, what do we have (..) just N-O-2?
Jan: I guess.
Sheila: Wait, should we start it? So, we start it and then this/ (*She points to the graphics window.*)
Jan: (*Jan reaches over and starts the simulation. The simulation proceeds illustrating the reversible conversion of NO_2 to N_2O_4.*)

Sheila: What HAPPENED? Something happened!
Jan: (*Speaking quickly*.) It made N-2-O-4.
Sheila: Ohhh! OK!
Jan: (*Jan records values onto her worksheet*.) So K is .958.
Sheila: Huh?
Jan: K is .958.
Sheila: Wait. Where are you getting that? That's time? Can we change that?
Jan: I don't know why it's like that (*She points to the monitor displaying the value of K*). I think it's/
Sheila: That's uh (…) OK. I am going to start it over again.
Jan: Let it go to 5 seconds.
Sheila: Ok, is this what we start out with?
Jan: Yeah, I think it's just...
Sheila: At this time (*5 seconds*), K is .95? (*She points to the K monitor. The simulation runs past the 5-second mark, and Jan reaches forward to stop it at 9 seconds*.)
Both: AWW!
Sheila: I think it skipped.
Jan: Can you just type in 5?
Sheila: No, we have to get the concentration back from time zero. Ummm (...)
Jan: Do we have the concentrations for those? (…) I think you have to start it to see it?
Sheila: So the concentration is .95 of this? (*She points to the K monitor*).
Jan: No that is K, which is/
Sheila: Huh? What's K? I think K is the concentration of this (*She points to the graphics window*).
Jan: No, K isn't concentration (…) It's like the point of equilibrium (…) It's like the constant of equilibrium. It/
Sheila: (*While Jan is explaining K, Sheila moves the pointer. When she moves the pointer over the plotting window, concentration values for each substance are displayed.*) No, wait, look you can just go here. (*She moves to a point on the plot were time is 5 seconds, and the value 38.3 appears, as in Figure 2a.*) Time at 5 is 38.3. (*They both write down 38.3 as the concentration of* NO_2.) That's about as close as it gets.

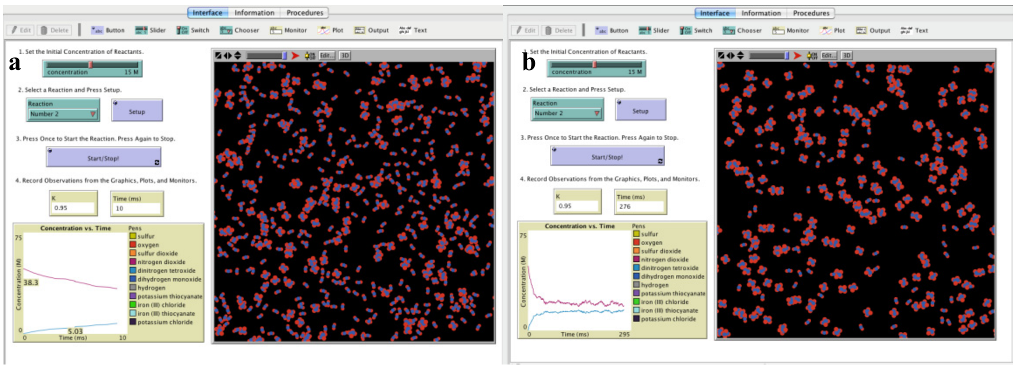


Figure 2. Screenshots of Reaction 1 at t = 9ms (a) and t = 275ms (b) When the Reaction Reaches Equilibrium.

In the above transcript, Sheila displays some confusion over where to locate important information in the interface. For example, Sheila appears unclear about what substance is represented by the graphics window and the meaning of 'K'. As she looks over the display, she mistakenly claims that the molecules in the graphics window are oxygen and that the value of K corresponds to the concentration of substances in the graphics window. In contrast, Jan immediately uses the key provided on the activity worksheet to state the molecular representation is NO_2 and the value of K is not concentration, but the equilibrium constant. In fact, the only information Jan cannot immediately locate in the display is the concentration. Gradually, Sheila comes to agree with Jan's explanation of the display and each representation, and she begins to investigate the interface on her own by moving the pointer over various windows. From her simple interaction, Jan discovers that she can locate the concentration of substances in the graphics window by placing the pointer over the relevant plot (Figure 2a). In this way, Sheila similarly supports Jan by helping her locate information necessary to complete the activity.

This simple four-minute interchange between Jan and Sheila suggests that their cooperative exchange allows them to come to understand each representation and feature of the display together. Although Sheila struggles at first to interpret the graphics display and monitors, she does not exclusively rely on Jan to interpret the simulation for her. Rather, with a small amount of guidance from Jan, she quickly apprehends the meaning of various elements in the display and in turn discovers a feature of the interface that allows her to locate information that Jan was seeking. As the girls proceed to complete the task, however, it becomes apparent that each makes a rapid initial decision about which representation, the macroscopic plot or the submicroscopic graphics, is most useful for determining if the reaction is reversible. Despite agreement between the girls on classifying the reaction as reversible, they at first ignore the evidence offered by the other.

Sheila: So, what is that (*She points to the graphics window*.)? N-O-2 is combining to form....

Jan: N-2-O-4 (*She points to the trace of N_2O_4 displayed on the plot.*)

Sheila: Was that in the first one though? Hmm (...) We don't really know/ So/ Do we have any of the reactants in there? (*She points to the graphics screen*.) Yeah, there's still reactants, N-O-2.

Jan: Well I think/ I don't/ we should let it run longer/ I want to see if (...) um (...) (*The girls are silent for ~10 seconds as they watch the simulation run.*) Is it reversible?

Sheila: No/ look/ Yeah, it's REVERSIBLE/ ‘cause look here (*She points to the graphics window*.) there's still N-O-2.

Jan: I know, but look they are getting smaller though (*She points to the plot of NO_2, which is gradually decreasing*.). I don't think it's reversible (...) (*After a few seconds, the simulation reaches a steady state and the plots of NO_2 and N_2O_4 stop changing*.) Oh wait/ Yeah, it's reversible. (*She points to the constant concentrations plots*.)

Sheila: (*Writing on her worksheet.*) Ok, so reversible. So, it's reversible and these little dudes are/ (*She points to the graphics window. Sheila notes the motion of the molecules on her observation and Jan records a sketch of the plot of her worksheet*.)

As Sheila and Jan record their observations of ‘Reaction 2’, it becomes evidence that each relies primarily on one representation to support the claim that the simulated reaction is reversible. Reasoning primarily from the submicroscopic representations in the graphics window, Sheila argues that because she can see that NO_2 still exists in the graphics window, the reaction is reversible. Indeed, her reasoning is sound as all of the NO_2 would convert into N_2O_4 were the reaction irreversible. At first, Jan disagrees with Sheila and argues that the reaction may be irreversible because the concentration of NO_2 is decreasing over time. She ignores Sheila’s appeal to the submicroscopic representation and insists they allow the reaction to run longer. Sheila urges Jan to “look here” at the graphics window to justify her claim that the reaction is reversible. Despite her appeal, Jan remains focused on the changing plot, and only agrees that the reaction is reversible when the concentrations become stable at equilibrium and the amount of NO_2 stops decreasing. Ignoring Jan’s appeal to the plots, Sheila records her observations of the submicroscopic window by noting the behavior of the “little dudes”, while Jan draws a sketch of the final plot shown in Figure 2b.

Now confident in their approach to the activity, the pair setup ‘Reaction 3’ (Figure 3a). In contrast to reaction two, ‘Reaction 3’ is an irreversible reaction that illustrates the condensation of gaseous oxygen and hydrogen to form liquid water. From their experience with ‘Reaction 2’, they are able to quickly locate the information in the display requested on the worksheet and rapidly fill in values. When deciding if the reaction is reversible, the girls disagree using information from different representations. As Jan and Sheila attempt to reconcile their disagreement over whether the reaction is reversible, they again show a bias toward using only one representation in the display. Moreover, each challenges the other’s interpretation of her chosen representation to establish an additional warrant for their different claims.

Jan: So, it's forming water/ basically forming water. (*She points to the plot of water concentration.*) So it must be irreversible if it's becoming constant (*She traces the constant value of water concentration on the plot with her finger.*)

Sheila: No, that's forming water. (*She points to the graphics window*.)

Jan: I know, but must be irreversible if it's staying constant. (*She points to the plot again*.)

Sheila: So then, what now do we put for the submicroscopic observations? Oh, it's REVERSIBLE.

Jan: (*She points to the plots of H_2 and O_2 which are decreasing logarithmically*.) Why do these keep going down, when this one (*She points to the plot of water.)* is constant?

Sheila: I don't know (…) Let's keep going. (*They watch as the simulation continues to run*.)

Jan: OH, it's becoming more and more water, it's becoming denser? (*She points to the bottom of the graphics screen where the water molecules have condensed.*

Sheila: Oh yeah. Well, look at that...

Jan: So, it's not/ it must be irreversible.

Sheila: No, it is reversible. Look! (*She waves her pen above toward the top of the graphics window were two* H_2 *and two* O_2 *molecules can still be seen as in Figure 3b.*)
Jan: But the/
Sheila: They are constant down there (*She points to the plots of* H_2 *and* O_2*, which are approaching 0 on the plot, and then she points again to molecules of hydrogen and oxygen that can be seen in the graphics window*.)
Jan: Probably not after a long time/ that (*She points to the two hydrogen and two oxygen molecules in the graphics window*.) is just because of evaporation.
Sheila: Oooooohhhh (….) Wait/ is^ it reversible?>

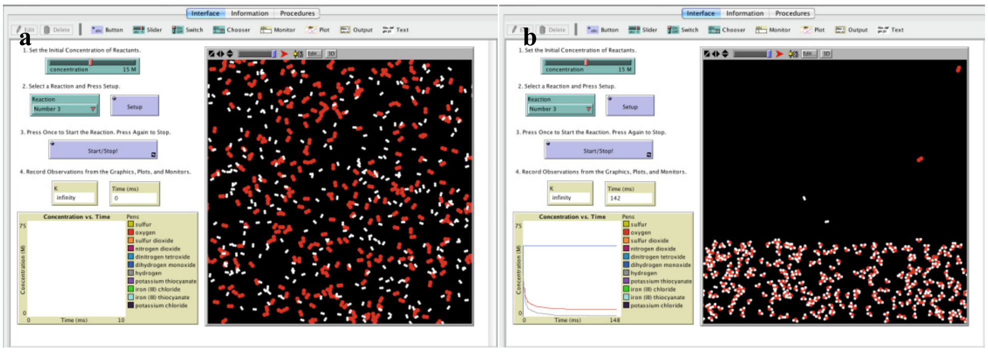


Figure 3. Screenshots of Reaction 3 at t = 9ms (a) and t = 142ms (b) When the Reaction Is Near Completion.

The above transcript illustrates how Jan again focuses on the constant concentration in the plot that suggests the reaction is irreversible; Sheila again focuses on the presence of starting reactants in the graphics window to justify her claim that the reaction is reversible. Interestingly, neither Jan nor Sheila appear confident and express confusion over what they are seeing. Namely, Jan cannot reconcile her observation that the concentration of the product (i.e., water) appears constant while the concentration of the reactants (i.e., hydrogen and oxygen) continues to decrease. Similarly, Sheila seems uncertain as to whether the reaction truly is reversible given only a few molecules of reactants can be seen and more reactants do not seem to be forming as in 'Reaction 2'. The girls' confusion is warranted: the reaction is irreversible, but they have stopped the simulation before it has run to completion. Thus, a few molecules of hydrogen and oxygen have yet to randomly collide to form water molecules and the graphics window shows them condensing into and evaporating out of the liquid water. Complicating the issue is that the concentration of liquid water is indeed constant and remains unchanging despite the formation of more water molecules as the reaction proceeds. Thus, the plot accurately shows the product concentration remains constant while the reactant concentration continually decreases.

Interestingly, the girls attempt to reinterpret each others chosen representation: Jan argues that Sheila's remaining reactants are due to "evaporation" and the reaction is therefore irreversible as reactants are not forming; Sheila points out that the hydrogen and oxygen concentrations have become "constant down there" and the reaction is therefore reversible because the concentration would otherwise decrease. As the pair continues to disagree over whether the reaction is irreversible, they call another student, Maria, over to settle the dispute. As the reaction continues, Sheila and Jan appear to 'trade' representations to garner additional evidence.

Jan: (*Calling over to Maria who is working with another student*.) Did you say reversible or irreversible because it's evaporating and then it goes back?
Maria: No, because it would have/ cause it would have broken apart again. (*She comes over to look at the screen.*)
Jan: But if you wait long enough this is what happens to it/ (*Jan starts the simulation and traces the constant concentration of water in the plot.*) and this. (*She points to graphics window and follows a hydrogen molecule as it travels across the screen. Two hydrogen molecules and two oxygen molecules can be seen on the screen, all others have reacted to form water, as in Figure 3b.*)
Maria: Nah, those eventually go back down.
Jan: Yeah, but more come back up, you can see it. (*She points to the graphics window*.)
Maria: (*In disbelief.*) Really!?^
Sheila: Hmmm…that looks pretty straight or does that mean it's reaching zero? (*She points to the plot. Jan and Maria lean forward and stare intently at the graphics window.)*
Maria: They will all go down eventually. (*Sheila looks over at the graphics window as two molecules dissolve into the water.*)
Sheila: Ah... *(Two molecules of oxygen evaporate out of the water*.) NOPE! (*All laugh*.)

Maria: I know^ it's IRREVERSIBLE! (*One hydrogen collides with an oxygen to form a water molecule. One hydrogen molecule and one oxygen molecule remain dissolved in the water and neither is discernable on the graphics screen.*)
All: OOOOHHHHH!
Sheila: There they go!
Maria: YES! (*The last hydrogen and oxygen molecules evaporate and become noticeable.*)
Jan: LOOK! ARE they going back up?
Maria: Huh? But, maybe those are just/ (The *molecules collide and form water*.)
Sheila: Ok! There we go!
Maria: SEE!
Jan: Ok, it is irreversible.

Jan explicitly states the problem to Maria: the evidence offered by each student, from the graphics window and the plot, is inconsistent. Although Maria is convinced that the reaction is irreversible, Jan and Sheila do not agree; they decide to run the simulation longer to gather more evidence for each respective claim. As the girls attempt to understand the reaction, an interesting shift in each student's source of evidence occurs. Jan focuses on the behavior of the molecules in the graphics window, which Sheila noted earlier. In contrast, Sheila focuses on the decreasing value of the hydrogen and oxygen traces on the plot, highlighted earlier by Jan. As the simulation runs, Sheila correctly notes that the concentration of hydrogen and oxygen are close to zero and not increasing; this is an important observation as it is strong evidence from the plot that the reaction is irreversible. Concurrently, Jan draws attention to the behavior of the four remaining reactant molecules to determine whether they are indeed interconverting between reactant and product or simply evaporating. The group's careful attention to the behavior of the molecules allows them to observe the moment when the hydrogen and oxygen molecules react to form (and remain) water molecules. In effect, each half of the pair was able to convince the other half to attend to information in a different representation and reconcile her observations to reach a consensus claim that the reaction is irreversible. Importantly, Sheila discredited Jan's evidence from the plot, and Sheila's evidence from the graphics window was discredited by Jan. As a unit, the dyad effectively and efficiently coordinated among the representations in the display to reach a conclusion.

Conclusion

The case study presented here offers two theoretical and methodological contributions to research on (meta)-representational competence with multi-representational displays. First, Jan and Sheila's collaborative coordination across representations in the display indicate that representational competence should not be conceived solely as a characteristic of individual students working with multiple representations. Rather, coordination across representations can occur through collaborative problem solving and pairs or groups of students can display representational competence that might go unnoticed in studies that focus on an individual student's behaviors alone. In the present work, the disagreements between Jan and Sheila led them to attend to information in multiple representations and support a shared claim. This finding is consistent with other research that has illustrated how groups coordinate representations (e.g., Stahl, et al., 2006; White & Pea, in press), and offers that coordination can occur when groups are working with one shared display that includes multiple representations. Thus, analyses of representational competence should include observations of students working together with multi-representational displays and avoid hasty claims about an individual's competence.

Second, each student's content knowledge and skill at interpreting representations can certainly provide a partial explanation for the observations. It is important to note, however, that in contrast to prior research on representational competence (e.g., Kozma & Russell, 1997; Russell, et al., 1997; Wu, et al., 2001), the students here displayed some basic competence interpreting, evaluating and coordinating across representations. While each student certainly displayed a bias for using one representation to support her claims, each attended to information present in other representations and attempted to reconcile conflicting interpretations. Of course, the underlying reason for each student's preference is not clear from the present analysis; certainly, there is no evidence that the bias is due to individual differences in media preference. Thus, as others have argued (e.g., Hammer, 1996; Stieff, et al., 2011), future research should focus on the competencies students bring with them to the classroom as opposed to skills that students lack.

Presently, the results of this study inform the design of activities using multi-representational displays. Perhaps the most obvious pedagogical implication is that curricular materials that incorporate multi-representational tools must include scaffolding to guide student attention to different representations. Indeed, this scaffolding needed might be more elementary than helping students to coordinate multiple representations: it must support students in attending to each displayed representation. As Sheila and Jan illustrate, individual students display preferences for using information in specific representations; working in isolation, each would need significant guidance in attending to non-preferred representations. However, the case illustrates that when working together from the same display, students are able to coordinate across representations to problem solve successfully. As seen here, inquiry activities that encourage students to provide multiple pieces of evidence from a multi-representational display to support their claims can foster representation coordination through planned argumentation. Simply put, the simultaneous displays of multi-representations do not afford 'discovery' of relevant information, nor are students easily able to use one representation to explain another. By devoting more attention in curricular materials to enhancing students' ability to attend to each displayed representation and coordinate between them *with their peers*, such technologies may result in the improved student understanding they are believed to promote in the science classroom.

Endnotes

(1) Transcription notations: / interruption or self-interruption, (.) one second pause, (*italics*) nonverbal actions, CAPS emphasis on word, ^ rising tone, > falling tone

References

diSessa, A. A., Hammer, D., Sherin, B. L., & Kolpakowski, T. (1991). Inventing graphing: Meta-representational expertise in children. *Journal of Mathematical Behavior, 10*(2), 117-160.

Hammer, D. (1996). Misconceptions or p-prims: How may alternative perspectives of cognitive structure influence instructional perceptions and intentions? *Journal of the Learning Sciences, 5*(2), 97-127.

Hoffmann, R., & Laszlo, P. (1991). Representation in chemistry. *Angewandte Chemie International Edition, 30*, 1-16.

Johnstone, A. H. (1993). The development of chemistry teaching. *Journal of Chemical Education, 70*(9), 701-705.

Kozma, R., & Russell, J. (1997). Multimedia and understanding: Expert and novice responses to different representations of chemical phenomena. *Journal of Research in Science Teaching, 34*(9), 949-968.

Kozma, R., Russell, J., Jones, T., Marx, N., & Davis, J. (1996). The use of multiple, linked representations to facilitate science understanding. In S. Vosniadou, R. Glaser, E. De Corte & H. Mandl (Eds.), *International perspectives on the psychological foundations of technology-based learning environments* (pp. 41-60). Hillsdale, NJ: Erlbaum.

Russell, J., Kozma, R., Jones, T., Wykof, J., Marx, N., & Davis, J. (1997). Use of simultaneous-synchronized macroscopic, microscopic, and symbolic representations to enhance the teaching and learning of chemical concepts. *Journal of Chemical Education, 74*(3), 330-334.

Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. pp. 409–426). Cambridge, UK: Cambridge University Press.

Stieff, M. (2005). Connected Chemistry–A novel modeling environment for the chemistry classroom. *Journal of Chemical Education, 82*(3), 489-493.

Stieff, M., Hegarty, M., & Deslongchamps, G. (2011). Coordinating multiple representations in scientific problem solving: Evidence from concurrent verbal and eye-tracking protocols. *Cognition and Instruction, 29*(1), 123-145.

Stieff, M., & Wilensky, U. (2003). Connected Chemistry - Incorporating interactive simulations into the chemistry classroom. *Journal of Science Education and Technology, 12*(3), 285-302.

White, T., & Pea, R. (in press). Distributed by design: On the promises and pitfalls of collaborative learning with multiple representations. *Journal of the Learning Sciences*.

Wilensky, U. (1999). NetLogo (Version 3.0) [Computer Program]. Evanston, IL: Center for Connected Learning and Computer Based Modeling, Northwestern University.

Wu, H.-k., Krajcik, J. S., & Soloway, E. (2001). Promoting conceptual understanding of chemical representations: Students' use of a visualization tool in the classroom. *Journal of Research in Science Teaching, 38*(7), 821-842.

Acknowledgments

The author gratefully acknowledges the support of Michelle McCombs and Comfort Ateh with data collection, which was supported by a grant from the Camille & Henry Dreyfus Foundation (SG-05-083). The Institute of Education Sciences, U.S. Department of Education, through Grant R305A100992 to the University of Illinois, Chicago, supported the research reported here. The opinions expressed are those of the author and do not represent views of the Institute or the U.S. Department of Education.

Guiding the Process of Argumentation: The Effects of Ontology and Collaboration

Frank Loll, Niels Pinkwart, Clausthal University of Technology, Clausthal-Zellerfeld, Germany
Email: frank.loll@tu-clausthal.de, niels.pinkwart@tu-clausthal.de

Abstract: Teaching argumentation is challenging, and the factors of how to effectively support the acquisition of argumentation skills through technology are not fully known yet. In this paper, we evaluate the impact of using an argumentation system with different argument ontologies and with collaborative vs. individual use on the outcomes of scientific argumentation. The results of a controlled lab study with 36 participants indicate that simple ontologies may be more appropriate than highly structured ones. In addition, collaborative argumentation lead to more cluttered argumentation maps, including a higher amount of erroneously used and duplicate elements, which indicates that an expected peer-reviewing between group members did not occur. Yet, groups also tended to include more points-of-view in their arguments, leading to more elaborated argument maps.

Introduction

The successful application of argumentation skills is important in many aspects of life. Even though this importance has been widely recognized, many people have problems as they engage in argumentation activities (Kuhn, 1991). In addition, educating students in their argumentation abilities is often not explicitly taught in school (Osborne, 2010) or at least problematic, caused (among other factors) by teacher's time: face-to-face tutoring is still the favored argumentation teaching method, but does not scale up well for larger groups.

One approach to deal with this issue is the use of argumentation systems (cf. Scheuer et al., 2010, for an overview) – tools that engage (groups of) students in argumentation by representing the argument in a graphical fashion (e.g., a graph, table or matrix), some of them providing feedback and intelligent support. While argument modeling has shown to be effective to promote learning in general (Harrell, 2007, Easterday et al., 2007), the specific roles of collaboration and of the argument ontology that is used in an argumentation system (i.e., the given palette of elements which "pre-structure" the input to the system) are still largely unclear and existing studies about their effects (e.g., Janssen et al., 2010; Osborne, 2010; Sampson & Clark, 2008; Schwarz & Glassner, 2007; Schwarz et al., 2000; Suthers, 2003; Toth et al., 2002) are hardly comparable because the specific conditions (population, the tool used, etc.) in which the studies were conducted differ a lot.

With respect to collaboration, it is widely accepted that arguing in groups can be beneficial for learning (e.g., Osborne, 2010; Schwarz et al., 2000). However, even though collaboration has shown to be effective for solving highly intellective problems (e.g., Laughlin et. al., 2006), reasonable collaboration does not occur by nature (Dillenbourg et al., 1995; Rummel & Spada, 2005) and unstructured collaborative argumentation per se will not lead to higher quality arguments (Sampson & Clark, 2008). Instead, it is important to aid the process of collaboration, e.g. though visualizations (Toth et al., 2003; Suthers, 2003, Schwarz & Glassner, 2007), access restrictions (Schwarz & Glassner, 2007) or collaboration scripts (Stegmann et al., 2007).

Ontologies are the elements available for modeling and representing an argument in the computer system. They may be as simple as in Athena (Rolf & Magnusson, 2002), where the elements are just node, pro and con, or as complex as in Rationale (van Gelder, 2007), a system with 6 categories for argument types and a large set of relations. Suthers (2003) highlighted the *representational guidance* of visualizations as well as of the ontologies that these visualizations refer to. However, even though it turned out in a couple of studies (e.g., Suthers (2003), Schwarz & Glassner (2007)) that the presence and the granularity of an ontology did indeed have an effect on the outcomes of argumentation, a concrete answer to the questions which argument elements are important in which domain (or to learn to argue in general) was not systematically investigated so far – Suthers (2003) just noted that a too detailed ontology may confuse students with "*a plethora of choices*" (p. 34).

Schwarz & Glassner (2007) investigated the effects of informal ontologies and floor control on the results of argumentation with respect to the quality of a final argument map (including the number of relevant claims and arguments, references to peers, and chat expressions) as well as to co-elaboration of knowledge. Their findings indicate that a higher structural degree of ontology and the presence of a turn-taking mechanism lead to more relevant claims and arguments and reduced unwanted behavior like off-topic discussions.

Stegmann et al. (2007) evaluated the usefulness of scripts that guide learners in creating single arguments by means of input masks that scaffold the creation of single arguments (consisting of claim, qualifiers and grounds) and a script that scaffolds a typical argumentation process consisting of argument-counterargument chains. The use of these scripts resulted in a gain in formal quality of single arguments as well as in argumentation sequences in online discussions. In addition, the scripts facilitated the acquisition of knowledge about argumentation.

Even though there is some evidence that these structural scaffolds may guide learners to successful interaction and success in learning, it is hard to come to general statements about their usefulness based on the existing studies with their multiple tools in various settings. In addition, possible interaction effects between collaboration and specific argument ontologies have not been investigated systematically: do some argument ontologies have benefits for collaborative usage as compared to others? In this paper, we want to make a next step towards a deeper, more comparable evaluation of the factors that make educational argumentation systems (un-)successful. First, we briefly outline a flexible framework that can be used to easily manipulate important variables. Then, we describe the use of this tool in a controlled lab study that investigates the impact of different ontologies and collaboration on the outcomes of scientific argumentation.

The LASAD System

The web-based LASAD framework for argumentation was designed with a focus on flexibility. Among other features (cf. Loll et al., 2010), the system is able to support different argument ontologies, visualizations, and collaboration settings. On the ontology side, each element, i.e. boxes and relations between them, can be independently defined based on a set of elements (e.g., labeled text fields, references to text passages or to URLs, ratings) including their visualization attributes such as color, size, or border. On the collaboration side, LASAD supports synchronous and asynchronous usage and offers tools like a chat, a list of users currently participating as well as awareness mechanisms such as the tracking of cursor movements of all participants. The use of these configuration options allowed us to set up the conditions for the study presented in this paper. A part of an argument map created via LASAD during the study presented in this paper is shown in Figure 1.

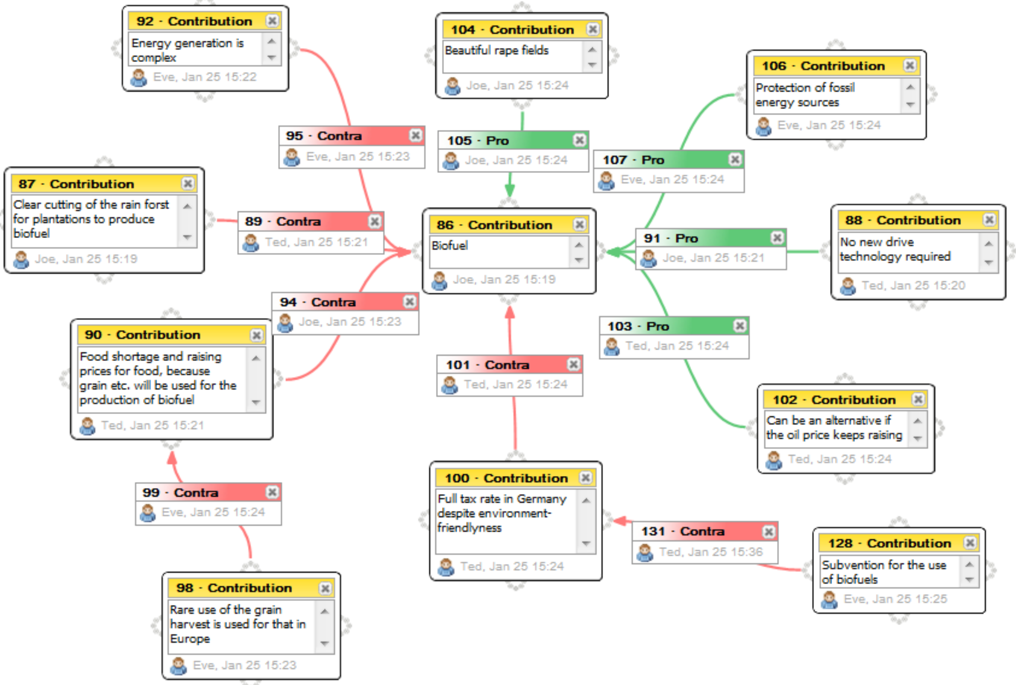


Figure 1. Part of an Argument Map created during the study via LASAD using a domain-independent ontology with general contributions showing arguments for and against biofuel.

Study Description

Inspired by prior research results on educational argumentation technologies (cf. above), our goal was to investigate the effects of using different ontologies in individual as well as collaborative argumentation on the outcomes of scientific argumentation and on the learning effects with respect to argumentation skills and domain-specific knowledge. Whereas prior research results suggest that argument models and ontologies may influence the process of argumentation (e.g., Suthers, 2003), we wanted to systematically evaluate the effects of various argument ontologies (all with graph-based visualizations) used for the same task. A second goal was to investigate how individual and collaborative argumentation differed, also in the context of different argument ontologies, to identify possibilities to aid argumentative processes better. More specifically, we were interested in the following hypotheses derived from prior research (cf. Scheuer et al., 2010):

Hypotheses: Effects of Collaboration

C1. Arguing in groups (as opposed to constructing arguments individually) will lead to a more elaborated argument, i.e. an argument of higher quality, due to different points-of-views of the participants.

C2. In collaborative argumentation activities, students will be more motivated than in individual ones. We hypothesize this based on the fact that discussions with other arguers will lead to a greater variation of the task steps and, hence, to a less monotonous activity. Prior results by Pinkwart et al. (2008) highlighted the importance of motivation to promote good learning results in argumentation activities.

C3. In collaborative sessions, the participation of single users may drop as compared to individual argumentation sessions: shy arguers may stop arguing against a dominant, leading group member.

C4. Collaborative argumentation will lead to more off-topic activities. Prior results by Schwarz & Glassner (2007) showed that groups can tend to get distracted from tasks, which can be detrimental for the overall argumentation process.

C5. Group members will review and respond to each other's arguments, and, hence, the overall number of mistakes will decrease in comparison with individual activities. We hypothesize this because argumentation is not a trivial undertaking: users may oversee their mistakes and, by discussing about parts of the argument, typical mistakes of single users may be revealed and corrected.

Hypotheses: Effects of Argument Ontologyies

O1. The higher the structural degree of an argument ontology is, the higher we expect the overall structure of the argument map to be. This is a direct consequence if the ontology is used correctly and supported by the findings of Schwarz & Glassner (2007).

O2. The more detailed an argument ontology is, the more elaborated the resulting argument will be. The rationale for this hypothesis is that we expect the multiple elements of detailed ontologies to prompt the users to make use of them and, hence, think about how to fill them with appropriate materials. Clark & Brennan (1991) also noted that it is easier to refer to knowledge units which have a visual manifestation, so that the presence of various, different ontology elements may lead to more discussions and, consequently, to a more detailed resulting argument.

Hypotheses: Interaction Effects

I1. For group argumentation, we hypothesize that the used ontology will influence the degree of collaboration: a more complex ontology may increase the need for collaboration (in order to discuss how to use the different elements to build an argument).

I2. In collaborative sessions, highly structured argument ontologies may be detrimental to the quality of the resulting argument (due to the double complexity of keeping track of the group process and using a complicated argument model at the same time), while the scaffolds that more structured ontologies provide may be more helpful in individual usage.

Study Design

To investigate the hypotheses, a mixed 3x2 design was used. The between-subject factor was the argument ontology. Here, the following three different ontologies were used:

1. A simple domain-independent ontology consisting of a general contribution type ("contribution") and three different relation types ("pro", "contra", "undefined").
2. A second, highly structured, domain-independent ontology based on the Toulmin argumentation scheme (Toulmin, 2003). It consists of five contributions ("datum", "conclusion", "warrant", "backing", "rebuttal") and four different relations ("qualifier", "on account of", "unless", "since").
3. A domain-specific ontology which was inspired by the Belvedere (Suthers, 2003) ontology, consisting of three contribution types ("hypothesis", "fact", "undefined") and three relation types ("pro", "contra", "undefined"). This ontology has shown to be effective for scientific argumentation.

The within-subject factor in the study was collaboration. Each participant was required to argue about one topic on his or her own and about another topic in a group of three. To eliminate possible confounds, we used counterbalancing so that half of the participants began with the group phase while the other half began with the single user phase. In the group phase, each participant worked on one machine. The participants were only allowed to communicate via the chat tool integrated in the argument framework. This simulated a remote discussion even though the users were located in the same room (the experimenter was in this room to enforce the rule). Overall, the study took 6 hours per user, including a 1 hour break between two sessions.

Tasks

Each participant worked on two open scientific problems that have no obvious solution. This kind of task choice was motivated by Toth et al. (2002), who used challenging science problems to simulate an authentic argument activity, avoiding a demotivation of students caused by hiding the answer of already solved questions. The Schwarz et al. (2000) results support this decision: their findings include that argumentation is most effective if students are arguing under uncertainty. In our study, the concrete topics for the arguments were:

1. The potential of alternative concepts for automotives (incl. the electronic car, the fuel cell, and biofuel)
2. The German energy mixture in 2030 (incl. nuclear power, fossil fuel, and renewable resources)

For each topic, three different possible positions were prepared. To allow all participants to argue for or against each of these positions, the students were provided with two pages of background information per position. This material, given in form of material chunks (graphs, tables as well as plain text) was typical for scientific argumentation, including facts, examples, statistical data and observations. In addition, there was one page containing material that was common to all positions (e.g., the definition of kilowatt hour for topic 2). The participants were allowed to go beyond the given material in their arguments.

Each session about a topic was split into four slots of 30 minutes each. In each of the first three slots, the participants were given the background material for one of the three positions (e.g., “nuclear power as a future energy”) as well as the common materials and were asked to create an argument about this position using the LASAD system. The fourth slot was the used to integrate the three separate positions and to draw a final conclusion to solve the argumentation task. For this last step, the participants were given the materials for all positions again.

Participants & Training

Overall, 36 (under) graduate students (25 male, 11 female) with different majors participated in the study. They were between 19-35 years old (m = 24.64, sd = 3.68) and in semesters 1 to 22 (m = 7.00, sd = 5.62). All participants were either native German speakers or fluent in this language (the complete study was conducted in German). Participation was voluntary and all participants were paid for completing the study. The participants were assigned randomly to all three “ontology” conditions, i.e. in each condition there were four groups consisting of three students each. In all but one group was one female student.

None of the participants had used the argumentation system before. Thus, a short video introduction (15 minutes) to the LASAD system was shown to make sure that all participants had the same basis. All videos consisted of three parts: (1) A general introduction how to interact with the system, (2) an overview of supporting features to work in groups (e.g., chat, cursor tracking), and (3) an ontology dependent part in which the condition dependent features of the system were explained using an example common to all conditions. Finally, the example argument that was presented in the video was distributed among all participants on paper and was available during the complete study.

Tests & Interviews

To test the learning effects caused by the argumentation tool’s use, three multiple-choice tests on argumentation abilities as well as two multiple-choice knowledge tests per topic were used. The tasks of the argumentation tests were taken from a list of questions of the Law School Admission Test (LSAT). Each argumentation ability test consisted of four questions, two from the area of logical reasoning and two from the area of analytical reasoning, i.e. we used approved questions that were not law specific. These tests took place before the first session, between the two sessions and after the second session. The order of the tests was counterbalanced. The participants were given 6 minutes (1.5 minutes per question) per argumentation test.

The knowledge tests were centered on the domain of argumentation in the respective study sessions (automotive concepts and energy mix). They were administered immediately before and after the corresponding sessions (in a counterbalanced manner) to measure domain learning. The participants were given 4 minutes (1 minute per multiple-choice question) per knowledge test.

In addition to these two tests, a questionnaire was used to evaluate the usability of the overall LASAD argumentation system. By means of this test, we wanted to check whether certain features of the system might have hindered the students to engage in reasonable argumentation, especially since this was the first larger study with the LASAD system. Here, the standardized *System Usability Scale* (Brooke, 1996) which has shown to be an accepted measure for usability (Bangor et al., 2009), was used.

Finally, we asked the participants in an open interview about their motivation during the study sessions, and about potential problems and ideas for future improvements of the system.

Coding Procedure

The *material* distributed to the participants consisted of unconnected information chunks including relevant as well as non-relevant parts. To be able to check how much of the relevant material was used, three domain experts independently created a list of all the facts that could either be directly taken from the material or directly concluded based on a combination of multiple information chunks. These lists were merged and discussed; the resulting lists (containing 81 entries for topic 1 and 75 for topic 2) were used as a reference for the relevant information that can be extracted from the hand-out material.

To get further insights into the resulting *argument maps*, 6 of 48 maps (one individual map and one collaborative map for each ontology, i.e. 12.5% of all the maps) were coded element-wise independently by two coders with respect to the use of given material. For each element (boxes and relations) in a diagram, the coders checked if the contained information was based on a fact in the "reference list" or if it was a completely new contribution. The coders also rated the correctness of the used ontology elements (if, for instance, a fact element was actually used to represent a fact). To judge the *structural quality of an argument map*, the coders additionally checked for each of the 6 chosen maps if this map contains (a) a starting hypothesis, (b) a conclusion and (c) a clear grouping of the different positions.

Based on these coding results, the inter-rater reliability was calculated and resulted in a Cohen's κ of 0.60 for the material used and 0.61 for the used elements. Concerning the general structural features (a-c), both coders agreed 100% on each measure. Taking into account the ill-defined nature of argumentation (Lynch et al., 2010), we assumed this level of agreement to be acceptable overall. The remaining elements were then coded by one coder in the same manner as described above. Overall, 5477 elements were manually coded this way.

To measure the *degree of coordination*, also the chat messages were encoded. First, the chats (consisting of 878 messages) were divided independently by two coders into episodes that belong together, e.g. a discussion about where to start with argument modeling. Slight differences were resolved by discussion between the coders. This resulted in an overall number of 196 chat episodes. Based on the chat episodes of three sessions (one per ontology, i.e. 25% of all material), the following four categories were agreed on as a coding scheme for the chat episodes: (1) *content*, (2) *structure*, (3) *coordination*, (4) *off-topic*. Based on this coding scheme, each chat episode within the 12 collaborative sessions was independently coded by two raters. The raters achieved a moderate Cohen's κ of 0.56. However, it turned out that the categories "*structure*" and "*co-ordination*" were often not clearly distinguishable so that these two categories were merged into one, which resulted in a high κ of 0.76. The raters resolved remaining conflicts through discussion.

Results

This study was the first one done with the LASAD framework. As such, we were also interested in the general usability of the system to check for any possible confounds related to weaknesses of the argumentation system we employed. The *System Usability Scale* test resulted in a mean score of 81.46 (which corresponds approximately to a "B" grade). The results thus indicate that the LASAD framework used in this study was perceived as an adequate tool to support argumentation.

Overall Effects on Argumentation Abilities and Domain Knowledge

Based on the scores of the argumentation ability tests ($m(T_1) = 1.611$, sd = 1.02; $m(T_2) = 2.056$, sd = 0.89; $m(T_3) = 2.083$, sd = 1.08; scale ranging from 0 to 4 points), a repeated measures ANOVA was calculated. This showed no statistically significant gains in argumentation skills, but a tendency ($F(2, 66) = 2.907$, $p = 0.062$). The between-subject factor "ontology" did not cause a significant effect ($F(2, 33) = 0.745$, $p = 0.483$).

Regarding the domain knowledge, a significant gain between pre/post-test scores was consistently achieved. In topic 1 (*Potential of Alternative Drive Concepts for Automotives*), the pre-test resulted in m = 0.92 (sd = 0.77), whereas the post-test resulted in m = 2.97 (sd = 0.88; based on paired samples t-test: $t(35) = -10.330$, $p < 0.001$; scale ranging from 0 to 4 points). In topic 2 (*The German Energy Mixture in 2030*) the pre-test resulted in m = 2.31 (sd = 1.04), whereas the post-test resulted in m = 3.42 (sd = 0.84; based on paired samples t-test: $t(35) = -5.976$, $p < 0.001$). Concerning the gain of domain knowledge, there was neither a significant difference between individual/collaborative use of the system nor between the different ontologies.

The Effects of Collaboration on the Argumentation Outcome

An ANOVA highlighted significant differences between individual and collaborative argument maps as shown in Table 1. In comparison, collaborative argument maps contained a larger amount of elements (i.e. boxes and relations between them) used overall ($F(1, 46) = 18.954$, $p < 0.001$) and a higher percentage of material used twice ($F(1, 46) = 6.983$, $p = 0.011$). Contrary to our expectations, the percentage of given material used did not differ significantly between individual and collaborative argumentation ($F(1, 46) = 0.932$, $p = 0.339$). Instead, group members provided significantly more *own* contributions (not derived from given material) ($F(1, 46) = 13.524$, $p < 0.001$) than individual arguers. Hypothesis C5 (groups will review the work of the members and, hence, will make less mistakes), measured by the percentage of wrongly used elements, has to be rejected ($F(1, 46) = 0.956$, $p = 0.333$). In fact, mistakes made in the group phases were often very similar to those made in the individual phases, e.g. wrong directions of relations. Thus, hypothesis C1 (group work → higher quality) is only partially supported. To measure the motivation of the participants, we analyzed the statements in the personal interviews conducted after the study. Here, all groups agreed (after short discussions) that working in groups was more motivating than working alone (hypothesis C2). This is supported by the observations of the experimenter, who stated that sometimes the participants in the individual sessions made a bored impression, as opposed to the collaborative sessions. Also, the groups always used all the time for their tasks, while some individuals finished early. Among our study participants, the question about the optimal group size for argumentation was discussed controversially. The majority agreed on two to three people arguing together;

larger groups and the resulting growing needs for coordination were seen as potentially detrimental for the overall results.

Table 1: Comparison between individual and collaborative argument maps.

	Individual (n=36)	Collaborative (n=12)
Overall # of used elements in the workspace	m = 104.00 (sd = 29.72)	m = 143.25 (sd = 15.80)
# of own contributions (not derived from given material)	m = 9.97 (sd = 5.43)	m = 18.58 (sd = 10.61)
Percentage of material used twice	m = 5.03% (sd = 4.23)	m = 9.14% (sd = 5.83)
Percentage of erroneous used elements	m = 22.46% (sd = 18.80)	m = 28.83% (sd = 21.73)

Hypothesis C3 (collaboration → participation drop of single users) is not easy to evaluate. We sought to investigate if users, when working together, became less active. To do so, we first computed the proportion of elements of each user in the collaborative sessions (min = 0.16, max = 0.59, m = 0.33, sd = 0.12) – i.e., single users created between 16% and 59% of a collaborative map. Apparently, there were thus no "drop-outs" and no dominating users creating the whole map alone. To represent how active a user is in individual sessions (as compared to his peers), we also computed, for each user, the proportion of his argument elements in his individual session to the sum of elements of all individual maps of his group members (min = 0.19, max = 0.51, m = 0.33, sd = 0.07). These two values resulted in a significant Pearson correlation of $\rho = 0.428$ ($p = 0.009$). Thus, hypothesis C3 can be rejected: users who are generally (in)active in individual sessions exhibit the same attitude also in collaborative sessions.

The hypothesis that working in groups might lead to a large amount of off-topic talk (hypothesis C4) could not be confirmed, as Table 2 shows. In the argument graphs, there were in fact no noteworthy off-topic contributions at all. The chat, embedded in the tool, seems to work quite well to avoid off-topic talk in the map.

Table 2: Overview of average chat episodes per ontology in multi-user maps.

Ontology	# Content Episodes	# Structure & Coordination Episodes	# Off-topic Episodes	Overall
m (Simple)	*6.25 (44.6%)*	*7.25 (51.8%)*	*0.5 (3.6%)*	*14*
m (Toulmin)	*5.00 (30.3%)*	*9.0 (54.5%)*	*2.5 (15.2%)*	*16.5*
m (Specific)	*4.75 (26.0%)*	*11.75 (64.4%)*	*1.75 (9.6%)*	*18.25*
m (Overall)	*5.33 (32.8%)*	*9.33 (57.4%)*	*1.58 (9.7%)*	*16.25*

The Effects of Ontology on the Argumentation Outcome

Based on the structural assessment of the maps (with respect to starting hypothesis, conclusion and clear grouping), no significant difference between different ontology conditions could be identified and, hence, hypothesis O1 (higher structural degree of ontology → improved structure of the argument) has to be rejected. However, users of the Toulmin-based ontology did show a tendency not to use a starting hypothesis ($F(2, 45) = 3.100$, $p = 0.055$), which is not really surprising as this ontology follows a different model of argumentation (beginning with data and then drawing a conclusion) and there is no explicit hypothesis element in the ontology.

Table 3: Overview of wrongly used ontology elements.

Ontology	Average percentage of wrongly used elements
Simple	m = 10.25% (sd = 10.80)
Toulmin	m = 41.29% (sd = 16.48)
Specific	m = 20.63% (sd = 16.57)

A difference between ontologies was found in the percentage of wrongly used elements, e.g., using a hypothesis box to represent a fact or to ignore the direction of a pro relation ($F(2, 45) = 18.082$, $p < 0.001$). A post-hoc Tukey HSD test indicated that there was a significantly higher error rate (shown in Table 3) in the Toulmin condition than in the others ($p < 0.001$ for Toulmin vs. Simple and $p < 0.001$ for Toulmin vs. Specific).

Hypothesis O2 (detailed ontology → elaborated arguments) could be confirmed partly. An ANOVA showed no significant differences ($F(2, 45) = 1.909$, $p = 0.160$) between ontologies with respect to the percent-age of given material being used. However, a non-parametric Kruskal-Wallis test indicated that the amount of own contributions (not derived from given material) did differ significantly ($p = 0.034$) as shown in Figure 2.

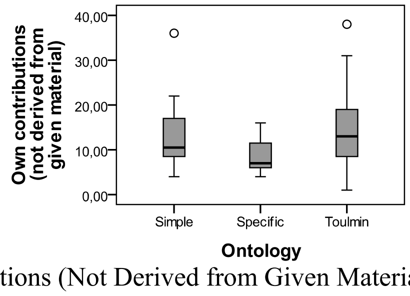


Figure 2. Differences of Own Contributions (Not Derived from Given Material) Used between Ontologies.

Interaction Effects

Regarding hypothesis I1 (ontology will influence the degree of collaboration), we analyzed the number of relations between elements of different authors in relation to the overall number of links as an indicator of the degree of collaboration (since this reflects the inter-relatedness of contributions from different users). An ANOVA did not reveal any significant difference between ontologies ($F(2, 9) = 1.689$, $p = 0.238$). Thus, the hypothesis could not be confirmed. Similarly, the results of a comparison of the number of chat messages used in different ontology conditions did not show any significant differences between ontologies as well (content episodes: $F(2, 9) = 0.212$, $p = 0.813$; structure & coordination episodes: $F(2, 9) = 0.408$, $p = 0.676$; off-topic episodes: $F(2, 9) = 0.568$, $p = 0.586$).

The comparison of the chat messages can be used for the investigation of hypothesis I2 (highly structured ontology will be detrimental to collaborative argumentation) as well, showing that the amount of needed coordination of structure and activities are not dependent on the complexity of the argument ontology. In addition, there was no significant interaction effect between individual / group argumentation and the ontology ($F(2, 42) = 0.605$, $p = 0.551$) in terms of the number of erroneously used elements for argumentation. As such, I2 has to be rejected.

Discussion

Regarding the knowledge and the argumentation tests, the results are not surprising: the increase of domain knowledge was an expected side-effect: if students argue about a topic for a longer time with additional material, the result that they have gained knowledge in this field can be expected. The positive trends shown by the argumentation ability tests is more interesting and needs to be further evaluated in long-term studies – 4 hours use of an argumentation system might not have been enough to come to significant effects at the .05 level.

With respect to collaboration, the results of our study confirm the possible benefit of collaboration for learning argumentation and are in line with prior findings (e.g., Janssen et al., 2010; Osborne, 2010; Sampson & Clark, 2008; Schwarz & Glassner, 2007; Schwarz et al., 2000). Against our hypothesis, groups in our study appeared not to have really checked each other's contributions well, but have argued for or against possible arguments, resulting in more elaborated arguments. This is clearly a point that may be worth future investigations as peer-reviews have shown to be an effective learning strategy (Cho & Schunn, 2007) and their inclusion into argumentation system could be fruitful. Based on a scripted approach, a peer-review process could be enforced in argumentation systems. Contrary to the results of Schwarz & Glassner (2007), the influence of structural aids and collaboration on the amount of off-topic talk could not be confirmed in our study. Possibly, the presence of a separate chat window was sufficient to keep the resulting argument map "clean".

Concerning the guiding function of the ontology, our results support Suthers' (2003) findings. The use of the Toulmin argumentation scheme did lead to a different style of argumentation: While the Toulmin approach is based on data used to draw a conclusion (without any hypotheses), the other ontologies used in our study employ hypotheses that are then backed up with supporting facts. However, we were *not* able to provide evidence that a domain-specific approach is more beneficial for the overall argument quality than a domain-independent one.

In addition, the participants in our study had problems with a highly structured argument ontology, confirming prior findings by Suthers (2003) that a broad range of elements may cause problems for students dealing with it: the Toulmin ontology puts excessive demands onto the students due to its complexity. In fact, there were even students who denied using the ontology correctly at all and only used the colors of the elements as orientation, e.g. using the red "on account of" relation as contra and the green "since" relation as pro. There was no noteworthy difference between the other two ontologies. Limiting, we would like to mention that the students were not familiar with any argument ontology before the study and the theoretical argument model of Toulmin was definitely the most complicated one in our study so that additional training may be required to deal with it. Also, a less elaborated ontology offers simply fewer possibilities to actually use elements incorrectly.

Conclusion & Outlook

In this paper, we contributed to the line of current research that investigates how to effectively support individual and collaborative argumentation. In the study reported in this paper, we systematically varied not only the role of collaboration and argument ontology on their own, but also looked at possible interaction effects between them. Our findings highlight the importance of adequate ontologies: the visual representation (and its complexity) of arguments makes a difference for the resulting overall argument quality. Yet, we were unable to find specific different needs of individuals vs. groups that would correspond to the employed ontologies. This could imply that groups are able to deal with quite complex ontologies even though they also have to manage the complexity of group work at the same time. Further, more detailed investigations are required here: one could argue that the structure provided by an ontology may even support the coordination in groups so that both effects (the detrimental one as well as the supportive one) may have canceled each other out. Finally, our results confirm that groups are able to enrich argumentation with different points-of-views.

In future research, we also plan to carry these results forward into other domains like ethics and legal argumentation and check if the found results are valid across argumentation domains. Additionally, we plan to compare scripted scenarios with unscripted ones to gain further insights how to improve the learning process.

References

Bangor, A., Kortum, P. T., & Miller, J. T. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. International Journal of Usability Studies, 4(3), 114–123.

Brooke, J. (1996). SUS - a quick and dirty usability scale. In P. W. Jordan, B. Thomas, & B. A. Weerdmeester (Eds.), Usability evaluation in industry (pp. 189–194). Taylor & Francis.

Cho, K. & Schunn, C. D. (2007). Scaffolded Writing and Rewriting in the Discipline: A Web-Based Reciprocal Peer Review System. Computers & Education, 48(3): 409-426.

Clark, H. H., & Brennan, S. E. (1991). Grounding in Communication. L. B. Resnick, J. M. Levine and S. D. Teasley (Eds.), Perspectives on Socially Shared Cognition (pp. 127-148).

Dillenbourg, P., Baker, M., Blaye, A., & O'Malley, C. (1995). The evolution of research on collaborative learning. In P. Reimann & H. Spada (Eds.), *Learning in humans and machines: Towards an interdisciplinary learning science* (pp. 189-211). Oxford: Elsevier/Pergamon.

Easterday, M. W., Aleven, V., & Scheines, R. (2007). 'Tis better to construct or to receive? Effect of diagrams on analysis of social policy. In R. Luckin, K. R. Koedinger & J. Greer (Eds.), Proceedings of the 13th International Conference on Artificial Intelligence in Education (pp. 93–100). IOS Press

Laughlin, P. R., Hatch, E., C., Silver, J. S., & Boh, L. (2006). Groups Perform Better Than the Best Individuals on Letters-to-Numbers Problems: Effects of Group Size. Journal of Personality and Social Psychology, 90(4): 644-651.

Loll, F., Pinkwart, N., Scheuer, O., & McLaren, B. M. (2010). Simplifying the Development of Argumentation Systems using a Configurable Platform. Educational Technologies for Teaching Argumentation Skills.

Lynch, C., Ashley, K. D., Pinkwart, N., Aleven, V. (2010). Concepts, Structures, and Goals: Redefining Ill-Definedness. International Journal of Artificial Intelligence in Education, 19(3): 253-266.

Harrell, M. (2007). Using Argument Diagramming Software to Teach Critical Thinking Skills. Proceedings of the 5th International Conf. on Education and Information Systems, Technologies and Applications.

Janssen, J., Erkens, G., Kirschner, P. A., & Kanselaar, G. (2010). Effects of Representational Guidance during Computer-Supported Collaborative Learning. Instructional Science, 38(1): 59-88.

Kuhn, D. (1991). The Skills of Argument. Cambridge University Press.

Osborne, J. (2010). Arguing to Learn in Science: The Role of Collaborative, Critical Discourse. Science, 328 (463): 463-466.

Pinkwart, N., Lynch, C., Ashley, K. D., & Aleven, V. (2008). Re-evaluating LARGO in the Classroom: Are Diagrams Better than Text for Teaching Argumentation Skills? In LNCS 5091 (p. 90 - 100).

Rolf, B., & Magnusson, C. (2002). Developing the art of argumentation. A software approach. Proceedings of the 5th International Conference on Argumentation (pp. 919–926).

Rummel, N. & Spada, H. (2005). Can people learn computer-mediated collaboration by following a script? In Fischer, F., Mandl, H., Haake, J. & Kollar, I. (Eds.), Scripting computer-supported communication of knowledge cognitive, computational, and educational perspectives. Dordrecht, NL: Kluwer.

Sampson, V., & Clark, D. (2008). The Impact of Collaboration on the Outcomes of Scientific Argumentation. Science Education, 93(3).

Scheuer, O., Loll, F., Pinkwart, N., & McLaren, B. M. (2010). Computer-Supported Argumentation: A Review of the State-of-the-Art. International Journal of CSCL, 5(1), 43-102.

Schwarz, B. B., & Glassner, A. (2007). The Role of Floor Control and of Ontology in Argumentative Activities with Discussion-based Tools. International Journal of CSCL, 2, 449-478.

Schwarz, B. B., Neuman, Y., & Biezunger, S. (2000). Two Wrongs May Make a Right… If They Argue Together! Cognition and Instruction, 18(4): 461-494.

Stegmann, K., Weinberger, A., & Fischer, F. (2007). Facilitating Argumentative Knowledge Construction with Computer-supported Collaboration Scripts. International Journal of CSCL, 2:421-447.

Suthers, D. D. (2003). Representational guidance for collaborative inquiry. In Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments, pp. 27–46.

Toth, E. E., Suthers, D. D., & Lesgold, A. M. (2002). "Mapping to Know": The Effects of Representational Guidance and Reflective Assessment on Scientific Inquiry. Science Education, 86(2).

Toulmin, S. E. (2003). The Uses of Argument. Cambridge University Press, 2nd rev. edition.

van Gelder, T. (2007). The rationale for Rationale. Law, Probability and Risk, 6(1–4), 23–42.

Acknowledgments

This work was supported by the German Research Foundation under the grant "LASAD – Learning to Argue: Generalized Support Across Domains".

The Role of Argumentation for Domain-Specific Knowledge Gains in Computer-Supported Collaborative Learning: A Meta-Analysis

Christof Wecker, Frank Fischer, Ludwig-Maximilians-Universität München, Leopoldstr. 13, 80802 München, Germany
Email: christof.wecker@psy.lmu.de, frank.fischer@psy.lmu.de

Abstract: The meta-analysis reported in this paper investigated the role of the quality of argumentation for domain-specific knowledge gains in computer-supported collaborative learning settings. Given the scarcity of primary studies that report correlations between these two variables, a meta-regression approach was used that uses interventions' effects on argumentation to predict their effects on domain-specific knowledge. Effect sizes for 17 comparisons extracted from 12 studies were included in the analysis using a random-effects model. On average, the interventions have a small to moderate effect on argumentation. With respect to the relation of their effects on argumentation to their effects on domain-specific knowledge, no unequivocal picture emerges. These findings call into question the broadly shared theoretical assumption that argumentation can be a mechanism that mediates the effects of interventions on domain-specific knowledge. A set of recommendations for strengthening future research on the topic is presented.

Introduction

It is a broadly shared conviction of CSCL researchers that argumentation can be a powerful mechanism to foster domain-specific knowledge (Andriessen, 2006; Osborne, 2010). Based on this assumption, a broad variety of tools and interventions to foster argumentation has been developed, ranging from direct instruction about characteristics of good argumentation (e. g. Choresh, Mevarech & Frank, 2009; Nussbaum, Sinatra & Poliquin, 2008; Yeh, K. H. & She, 2010) to argumentation maps (e. g. Janssen, Erkens, Kirschner & Kanselaar, 2010; Munneke, van Amelsvoort & Andriessen, 2003; van Drie, van Boxtel, Jaspers & Kanselaar, 2005; van Drie, van Boxtel, Erkens & Kanselaar, 2005; Schwarz, Neumann, Gil, J. & Ilya, 2003) and from different kinds of discussion seeds (Clark, D'Angelo & Menekse, 2009) to collaboration scripts (Kollar, Fischer & Slotta, 2007; Stegmann, Weinberger & Fischer, 2007; Weinberger, Stegmann & Fischer, 2010). These tools have been used for a while now in order to induce interactions among learners that are characterized by high argumentative quality and to study the role of argumentation for the learning of domain-specific content. Although some of these studies employ a qualitative methodology and thereby provide detailed accounts of the mechanisms involved in learning through argumentation, also a considerable number of quantitative studies on the topic have been conducted.

Therefore, it seems to be the right time now for integrating these findings using quantitative methods of research synthesis in order to evaluate the evidence pertinent to the widespread convictions about the role of argumentation for learning. On closer inspection, however, many of the studies focus only on some of the variables that are relevant for the issue. For instance, sometimes only effects of an intervention on the quality of argumentation during interaction in a collaborative learning phase (e. g. Munneke et al., 2003) or on the acquisition of argumentation skills as measured by a posttest are measured and reported. Among the studies that measure both the quality of argumentation during collaboration and learning outcomes concerning domain-specific knowledge, only a small proportion either reports correlations between argumentation during learning and domain-specific knowledge acquisition (e. g. Choresh et al., 2009) or employs similar analytical strategies that allow for an assessment of the association between these two kinds of variables (e. g. Asterhan, 2008). Often, only correlations within some of the several experimental conditions are reported along with the information that no significant correlation was found in the other conditions (e. g. van Drie et al.,2005a).

This practice makes the conventional approach of integrating indicators of association between the two variables of interest unfeasible. As a sufficient number of the studies in CSCL research report effects of interventions on both measures of argumentation and measures of domain-specific knowledge, a different analytical approach is used in this meta-analysis: If the assumption about the role of argumentation for domain-specific knowledge acquisition described above is true, interventions that produce strong positive effects on argumentation should also produce strong positive effects on domain-specific knowledge, and interventions that produce small or even negative effects on argumentation should produce small or negative effects on domain-specific knowledge. Therefore, the technique of meta-regression is used in a somewhat unusual way to predict interventions' effects on domain-specific knowledge on the basis of their effects on argumentation in order to test the validity and generality of the importance of argumentation for domain-specific learning. This approach presupposes substantial variation in effect sizes with respect to both variables. Given the diversity of argumentation interventions used in primary studies, it seems likely that effect sizes are actually quite heterogeneous. To test whether this prerequisite is fulfilled, in addition to the covariation of effects on argumentation and domain-specific knowledge also the magnitude and variation of effects of the interventions on these variables are scrutinized.

Research Questions

This meta-analysis addressed the following research questions:

(1) What is the mean effect of argumentation interventions on argumentation, and are the effects of the single argumentation interventions homogeneous? It is expected that on average the argumentation interventions have a positive effect on argumentation. As the kinds of argumentation interventions investigated in CSCL research are quite different, systematic variation of their effects on argumentation is expected.

(2) What is the mean effect of argumentation interventions on domain-specific knowledge, and are the effects of the single argumentation interventions homogeneous? According to the assumption that argumentation mediates intervention effects on domain-specific knowledge, on average the argumentation interventions should have a positive effect on domain-specific knowledge. As in the case of effects on argumentation, systematic variation of effects on domain-specific knowledge is expected due to the heterogeneity of argumentation interventions studied in CSCL research.

(3) How are argumentation interventions' effects on argumentation related to their effects on domain-specific knowledge? According to the theoretical assumptions about the role of argumentation for domain-specific knowledge acquisition, a moderate positive relation between effects on argumentation and effects on domain-specific knowledge is to be expected.

Method

Selection of Studies

Criteria for Inclusion

The purpose of this meta-analysis is to provide generalized information about the relation of the quality of argumentation during learning to domain-specific knowledge in a setting of computer-supported collaborative learning. Therefore, the criteria for the inclusion of a study were the following:

(1a) The quality of argumentation in collaboration during the learning phase or – as a proxy indicator for this variable – the learners' acquired argumentation skills after the learning phase was measured.

(1b) The individual learners' domain-specific knowledge about the content discussed during the learning phase was measured.

(2) The study was conducted in a computer-supported collaborative learning setting.

(3) At least one effect of the intervention under investigation on the quality of argumentation or on argumentation skills on the one hand and at least one effect on individual domain-specific knowledge on the other hand as well as their associated variances can be determined based on the information provided.

Search Strategies

The approach to locate the studies that fulfill these criteria comprised several complementary strategies. First, the bibliographic databases ERIC and PsycINFO were searched. The search terms "argument" (truncated), "learning" (truncated) and "CSCL" (or a conjunction of several synonyms of "collaboration" (truncated) and "computer") were used conjunctively. These search terms were not limited to specific fields to warrant high recall at the cost of lower precision. This yielded 370 publications many of which were either unrelated to the topic or did not report empirical studies. Based on a thorough review of the abstracts of these publications, each publication for which it could not be ruled out definitively that they might fulfill the criteria for inclusion were selected for inspection. Second, relevant studies mentioned in the reference lists of studies and reviews retrieved were also included in the set of studies for integration. Finally, the digital versions of the proceedings of recent CSCL and ICLS conferences were searched electronically for occurrences of "argument" to yield further studies on the topic that have not been published in journals yet.

Sample of Studies

The sample included in this meta-analysis comprises 12 studies. The integration of effect sizes ís based on data from more than 1400 persons.

Coding of Variables

Two types of variables coded from the primary studies were included in this meta-analysis: Each study's effect on argumentation and each study's effect on domain-specific knowledge. The types of intervention used in the primary studies were also coded. As the focus of this paper is not on the effects specific or diverse kinds of interventions on the association between argumentation and domain-specific knowledge gains, characteristics of interventions and similar study features are not used in the analyses presented here.

Argumentation

All variables in a study that were either indicators of specific aspects of the quality of argumentation in collaboration during the learning phase or indicators of the learners' acquired argumentation skills after the learning phase were selected for coding. The types of variables included comprised indicators of the number of (specific types or components of) arguments as well as their argumentative or content-related quality on a micro level of analysis and the number and quality of specific argumentative speech acts on a higher level of analysis. If a study contained several types of measures of argumentation, the set of most proximal ones were selected. For example, if both indicators of argumentation skill in a posttest and argumentative quality of contributions during the learning phase were reported, only the latter were included in the analysis. If indicators of both argumentative strength and content-related accuracy of contributions during the learning phase were provided, only the former were included. The rationale behind this approach was to use the best estimates available of the effect of an intervention on the argumentative quality of contributions during the learning phase, which according to the theoretical assumptions should mediate an intervention's effect on domain-specific knowledge.

For each of these variables, an effect size and its variance were computed from the descriptive statistics (means, standard deviations and sample sizes of subgroups) provided in the study. If part of this information was insufficient, either any kind of measure of effect size – if reported in the studies – was extracted directly and transformed, if necessary, or the effect size was computed from the values of inferential test statistics.

Domain-specific Knowledge

All variables in a study that quantified individually attributable domain-specific knowledge due to the learning experience were coded. Individual attribution means that (in one case) conceptual change diagnosed for individual participants on the basis of discourse data from an online discussion was included, whereas the domain-specific quality of essays written by groups could not be used as an indicator of individual domain-specific knowledge. In the majority of cases this kind of data was extracted from posttests. Effect sizes were determined as in the case of argumentation.

Calculation and Statistical Analysis of Effect Sizes

Effect Size Metric

As most of the studies report main effects from between-subjects experimental designs, the unbiased estimator d of standardized mean differences suggested by Hedges and Olkin (1985, p. 81) was used for integrating the study effects. If pretest data were available, posttest effect sizes were corrected by subtracting the corresponding pretest effect size from the posttest effect size.

Statistically Dependent Effect Sizes

If for a study separate effect sizes could be determined for different subsamples (such as the different values of a second factor in a 2x2 design that is not relevant in the present context), the data from all subsamples were included as independent effect sizes (cf. Borenstein et al., 2009, p. 218 f.).

Other studies (typically ones that use a one-factorial design) compared several treatment groups to a common control group. In this case, effect sizes from the same study are correlated due to the values from the common control group that enters the effect sizes of all experimental groups (Borenstein et al., 2009, p. 239-241). One option to deal with this kind of dependency is to compute one common average effect size for the different intervention groups and use only this single effect size for further integration. As the purpose of the present meta-analysis is to determine the association between argumentation and domain-specific knowledge gains on the study level and averaging across intervention types would blur relevant variation and covariation, the effect sizes of the single comparisons from this type of study are used separately in the integration. We regard the partial violation of assumptions of the statistical methods employed for integration as less severe than the loss of important information we would have to suffer otherwise.

Finally, most studies report findings for a whole set of indicators of the quality of argumentation, and some also report findings for several scales of domain-specific knowledge or for multiple measurement time-points using the same instruments. In these cases, the effect size indices from each comparison were aggregated to yield one effect size for argumentation and one effect size for domain-specific knowledge per comparison. As most studies do not report correlations between the different dependent measures, we ran all analyses reported in this paper with three different global assumptions about the covariation among dependent measures ($r = 0$; $r = .3$ and $r = 1$). There were no differences between the three variants in any of the analyses that would lead to a different answer to any of our research questions. Therefore we consistently report only the results from the analyses assuming independence of multiple measurements.

These decisions resulted in the integration of 17 pairs of separate effect sizes for argumentation and for domain-specific knowledge. They were calculated on the basis of 55 original effect sizes extracted from the primary studies. For 11 comparisons we could extract effect sizes for argumentative quality of contributions, whereas for 6 comparisons we had to rely on the more distal posttest measures of argumentation skill.

Method of Integration

The primary studies included in this meta-analysis investigated the effects of a very broad array of argumentation interventions that ranged from instruction about good argumentation to the use of different kinds of discussion seeds and from argumentation maps to collaboration scripts. Therefore no unique common effect size can be assumed for these studies. Accordingly, the random-effects approach for integrating the findings was employed. The between-studies variance component was estimated using the method described by Raudenbush (1994, p. 310 f.).

Results

Research Question 1: Effects of the Argumentation Interventions on Argumentation

The individual effect sizes for argumentation and their 90 % confidence intervals as well as their summary effect are presented in figure 1.

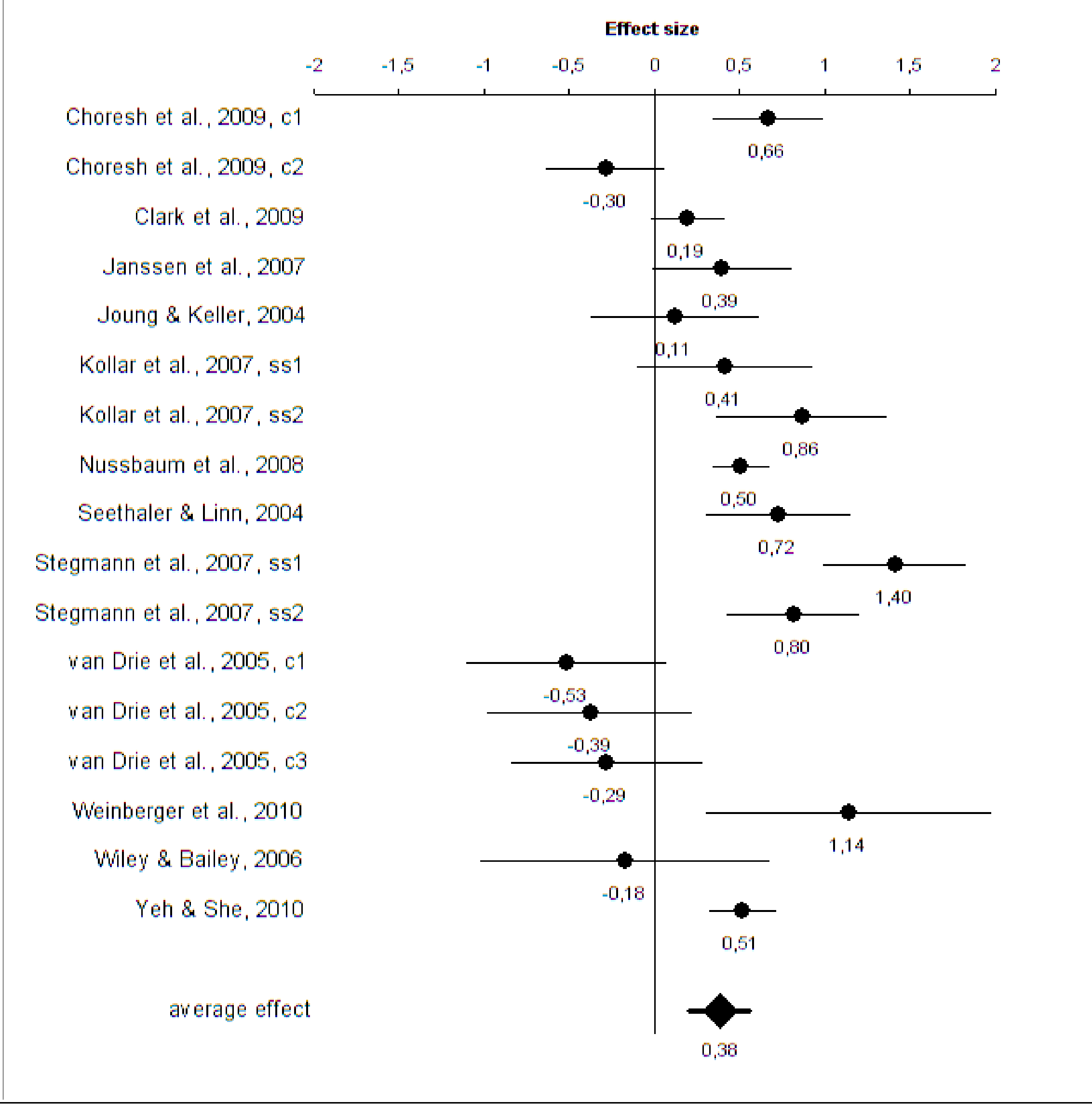


Figure 1. Forest plot of the interventions' effects on argumentation.
The error bars indicate 90 % confidence intervals. "c" along with a number refers to multiple comparisons with a common control group, "ss" along with a number refers to independent sub-samples.

A small to moderate average effect size of 0.38 was estimated, $SE = 0.11$; $CI_{90\%} = [0.20; 0.56]$; $p < .01$, one-tailed. The test for the homogeneity of the effects of the interventions on argumentation indicated variation in the true effect sizes, $Q_T(df = 16) = 58.99$; $p < .01$; $I^2 = .73$.

We further conducted a moderator analysis comparing the group of effect sizes based on measures for argumentative quality of contributions ($d = 0.32$; $SE = 0.15$; $CI_{90\%} = [0.07; 0.57]$; $p = .02$, one-tailed.) with the group based on the more distal posttest measures of argumentation skill ($d = 0.46$; $SE = 0.16$; $CI_{90\%} = [-0.19; 0.73]$; $p < .01$, one-tailed.). Effects from these two groups were homogeneous, $Q_B(df = 1) = 0.40$; $p > .99$.

These findings provide support for the validity of our approach: First, because there is substantial variation in effects on argumentation, which is required for detecting the association of interest, it is possible to investigate the role of argumentation for domain-specific knowledge gains on the level of studies. Furthermore, the inclusion of measures of post-treatment argumentation skill as more distal indicators of argumentative quality of contributions seems warranted given the homogeneity of effect sizes of both types. However, because the focus of the present analysis is not on the effects of different kinds of intervention on argumentation, no further attempt to explain the variance in effects on argumentation is made in the following.

Research Question 2: Effects of the Argumentation Interventions on Domain-Specific Knowledge

On average, argumentation interventions have no effect on domain-specific knowledge, $d = -0.04$; $SE = 0.14$; $CI_{90\%} = [-0.28; 0.19]$; $p = .62$, one-tailed. However, the test for the homogeneity of the effects of the interventions on domain-specific knowledge indicates variation in the true effect sizes, $Q_T(df = 16) = 129.46$; $p < .01$; $I^2 = .88$. To study the role of argumentation for domain-specific knowledge gains, this variation is further analyzed as a function of the interventions' effects on argumentation in the following.

It should be noted, however, that in one particular study large negative effects on domain-specific knowledge were observed (van Drie et al., 2005a). These effects were more than two standard deviations lower than the average effect size of the other studies in the sample, which means that they fulfil the frequently applied criterion for the exclusion of outliers. If the three effect sizes are excluded from the analysis, the estimate of the average effect size on domain-specific knowledge changes to $d = 0.17$; $SE = 0.13$; $CI_{90\%} = [-0.05; 0.39]$; $p = .10$, one-tailed. But also without these three effect sizes the test for the homogeneity of the effects still indicates variation in the true effect sizes, $Q_T(df = 16) = 84.74$; $p < .01$; $I^2 = .85$.

Research Question 3: Association between the Effects of the Argumentation Interventions on Argumentation and their Effects on Domain-Specific Knowledge

A two-dimensional forest plot (which is a variant of a scatterplot with some additional information) of the individual effects on argumentation and domain-specific knowledge is displayed in figure 2.

To determine the association between the interventions' effects on argumentation and on domain-specific knowledge, a meta-regression with the effects on domain-specific knowledge as the criterion and the effects on argumentation as the single predictor was conducted. A positive estimate for the unstandardized regression coefficient was obtained, $b = 0.58$; $SE = 0.31$; $CI_{90\%} = [0.06; 1.10]$; $\beta = .44$; $p = .03$, one-tailed. The higher the effects of the interventions investigated in the primary studies on argumentation, the higher were also their effects on domain-specific knowledge.

If the three comparisons from the van Drie et al. (2005a) study are excluded as outliers on the domain-specific knowledge dimension, however, an association of argumentation and domain-specific knowledge can no longer be detected, $b = -0.12$; $SE = 0.30$; $CI_{90\%} = [-0. 61; 0.37]$; $\beta = -.12$; $p = .34$, one-tailed.

Discussion

The results of this meta-analysis provide evidence that on average argumentation interventions developed in CSCL research are successful with respect to their most proximal goal of enhancing argumentation. As expected, these interventions do not share one common effect size. This is an advantage for the goal of the present integration to investigate the role of argumentation for domain-specific knowledge acquisition on the study level, given the scarcity of correlations on the person level reported in primary studies. However, differential effects of different types of interventions on argumentation should also be investigated meta-analytically in the future.

This broad variety of effects on argumentation is accompanied by an average zero effect of these interventions on domain-specific knowledge on the basis of, again, varying true effects. Even if the van Drie et al. (2005a) study is removed as an outlier, no significant effect on domain-specific knowledge emerges. Given the broad consensus among CSCL researchers about the role of argumentation for learning, it is quite remarkable that on average argumentation interventions do not produce beneficial effects on domain-specific knowledge acquisition.

The picture becomes even more obscure when considering the relation between argumentation and domain-specific knowledge: Only if the van Drie et al. (2005a) study is retained in the sample of studies are the effects of the interventions on argumentation positively related to their effects on domain-specific knowledge. If this study is removed because it qualifies as an outlier with respect to domain-specific knowledge, no such relation can be detected. On inspection, however, the kinds of intervention investigated in this study do not stand out from the ones used in other studies in the field in any obvious way.

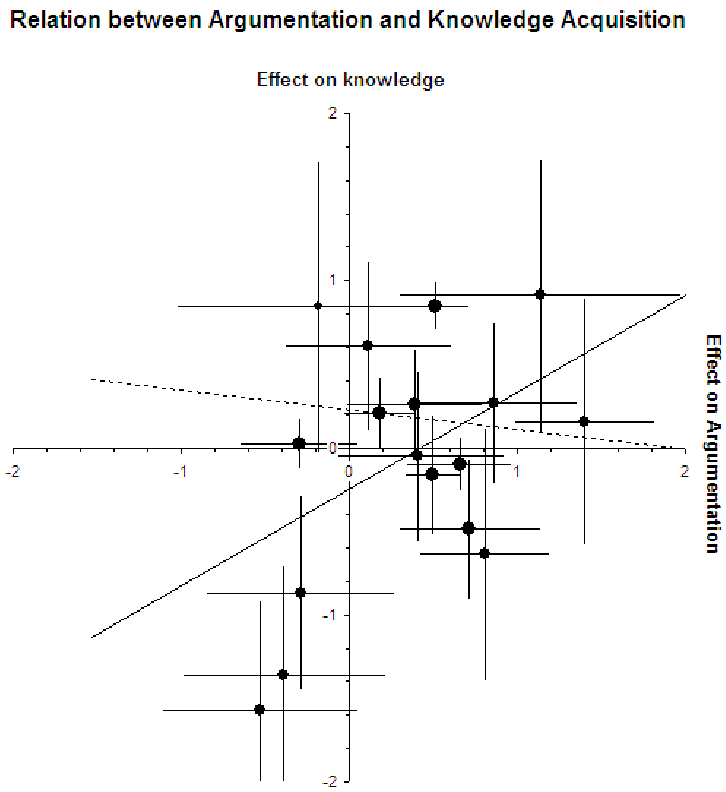


Figure 2. Two-dimensional forest plot of the interventions' effects on argumentation and individual domain-specific knowledge.

Effects on argumentation are displayed on the x-axis, effects on domain-specific knowledge on the y-axis. Each dot represents the ("two-dimensional") effect size for domain-specific knowledge and argumentation for one comparison from the primary studies. The size of each dot is proportional to the comparison's weight according to the random effects model. 90 % confidence intervals are displayed by the horizontal lines for effects on argumentation and by vertical lines for effects on domain-specific knowledge. The continuous diagonal line represents the equation of the meta-regression including all studies ($d_{knowl\_i} = -0.249 + 0.580 \cdot d_{arg\_i}$), whereas the dotted line represents the meta-regression after exclusion of the van Drie et al. (2005a) study ($d_{knowl\_i} = 0.227 - 0.120 \cdot d_{arg\_i}$).

Before discussing the consequences of these findings, potential limitations of this meta-analysis have to be taken into account. Its main shortcoming is its reliance on meta-regression to predict study-level or comparison-level effects on domain-specific knowledge on the basis of effects on argumentation instead of the more common approach of integrating within-study correlations between argumentation and domain-specific knowledge gains. However, apart from the fact that the latter is currently not feasible, the former approach constitutes a stringent test of the assumption that argumentation fosters domain-specific knowledge acquisition. While a correlation between effects of argumentation interventions on argumentation and effects on domain-specific knowledge does not imply that argumentation quality and domain-specific knowledge acquisition are correlated on the level of groups or individuals, the lack of a correlation on the level of experimental comparisons provides some indication against the assumption that argumentation fosters domain-specific knowledge acquisition, at least as they were measured in the studies included in the analysis.

What can the findings from this meta-analysis tell us? First of all, the body of research that has addressed the role of argumentation for domain-specific knowledge acquisition in CSCL using a quantitative methodology that allows for meta-analytic integration, is not abundant. Given this relative scarcity of evidence, more caution with respect to general statements about the effects of argumentation on domain-specific learning seems advisable (e. g. Osborne, 2010, p. 464 f.).

Furthermore, as this meta-analysis shows, the quantitative evidence amenable to integration currently does not provide the compelling picture one would expect, given the dominating view in the field. Rather it suggests reconsidering and even questioning cherished beliefs. It can be questioned, just to give an example, that arguments are "better", i. e. more functional for domain-specific knowledge acquisition, if more parts of them are mentioned explicitly, which is used as an indicator of argumentation quality in several coding schemes (cf. Clark, Sampson, Erkens & Weinberger, 2007) some of which were applied in the studies in the present meta-analysis. A more promising approach could be the analysis of argumentation quality based on a typology of argument schemes that are appropriate for different types of claims (cf. Walton, Reed & Macagno, 2008) because this approach links the formal quality of arguments closer to the content. Another assumption that might be called into question is what might be called the "collaborative enrichment" assumption that argumentation is the joint elaboration of content, which fosters individual knowledge acquisition (e. g.. Andriessen, 2006, p. 445). The mechanisms by which argumentation may sometimes provoke the dismissal of deeply entrenched views and thereby advance understanding, might be slightly more intricate.

Certainly the present meta-analysis does not force us to jettison the conviction that argumentation can lead to domain-specific learning. It should persuade us, however, to develop more accurate accounts of both argumentation quality and the mechanisms by which good arguments in a discussion among learners may influence the further progression of the discussion as well as the learners' cognitive processing and thereby lead to understanding of the content. Therefore, we suggest that future research should address the following points that are located on a theoretical level:

(1) As already indicated, in general, more thorough theoretical accounts of mechanisms that might explain beneficial effects of argumentation on domain-specific knowledge are needed.

(2) Based on such assumptions, a comprehensive taxonomy of any relevant aspects of the amount and quality of argumentation could be developed that allows comparing and integrating studies investigating the same aspects of argumentation. Such a taxonomy could differentiate, among other things, the occurrence or number of different types of arguments, their average or aggregated tenability and relevance, the breadth and depth of topics covered in the discussion (Munneke et al., 2003) or the occurrence of specific argumentative speech acts.

In terms of methodology, based on our experiences, we urge researchers in the field to consider the following points when reporting empirical research about the relation of argumentation and learning in CSCL:

(3) Fine-grained analyses of patterns of argumentation and their relation to learning as they are often found in qualitative studies of the same topic should be translated into quantitative indicators that allow for integration and thereby an assessment of the generality of the findings. Sample size is not so much of an issue in this respect, as evidenced by a study by Wiley and Bailey (2006) with only 8 participants: This study could be integrated in this meta-analysis because the analyses were quantified and all necessary information was reported. This recommendation has the goal to make the often more informative and more valid analyses of argumentation during collaboration typically found in qualitative studies (instead of the more distal acquisition of argumentation skills as measured by posttests) accessible to integration.

(4) Furthermore, it deserves mention that all statistics necessary to compute effect sizes need to be reported in publications. As a general rule, descriptive statistics (including standard deviations and subsample sizes) should be provided. Exact p-values and effect sizes should be presented (cf. APA, 2001, p. 138 f.), even in the case of insignificant effects, to allow for unbiased integration. If several measures for the same general variable are used in a primary study, they often need to be collapsed in a meta-analytical integration. This is strongly facilitated if the intercorrelations among the separated indicators are known. Therefore it is desirable that correlation matrices for sets of cognate indicators are presented, as they can often be found in research using multi-dimensional questionnaires. Furthermore, a more thorough practice of reporting indices of intercoder objectivity and reliability would allow for corrections for the attenuation of relationships between variables that is due to error variance in measurements. Most importantly, these indicators should not be averaged across variables, but reported separately for each individual variable used in the statistical analyses (cf. De Wewer, Schellens, Valcke & Van Keer, 2006, p. 11).

(5) As a complement to the current meta-regression approach it is still desirable to integrate findings about person-level covariation between argumentation and domain-specific knowledge acquisition. For this purpose, correlation coefficients are needed that quantify the association between argumentation during the learning phase and domain-specific knowledge as measured in post-tests. To avoid over-estimation of the importance of argumentation for learning, the calculation and presentation of partial correlations controlling for individual motivational and cognitive learning prerequisites such as general cognitive abilities or – as a minimum requirement – domain-specific prior knowledge on the basis of the whole sample rather than specific experimental conditions should become a common standard.

We believe that if these suggestions are picked up in future research, we will soon know more about the role of argumentation for computer-supported collaborative learning than we know now.

References

References marked by an asterisk (*) indicate studies that were included in the meta-analysis.

Andriessen, J. (2006). Arguing to learn. In R. K. Sawyer (Hrsg.), *Cambridge handbook of the Learning Sciences* (pp. 443-459). Cambridge: Cambridge University Press.

American Psychological Association (APA). (2003). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: American Psychological Association.

Asterhan, C. S. C. (2008). Processes of argumentation and explanation in conceptual change: Results from protocol analyses of peer-to-peer dialogue. In ISLS (Ed.), *International perspectives in the Learning Sciences: Cre8ing a learning world. Proceedings of the Eighth International Conference for the Learning Sciences - ICLS 2008* (vol. 1, pp. 60-67). ISLS.

Borenstein, M., Hedges, L. V., Higgins, J. P. T. & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester: Wiley.

*Clark, D. B., D'Angelo, C. M. & Menekse, M. (2009). Initial structuring of online discussions to improve learning and argumentation: Incorporating students' own explanations as seed comments versus an augmented-preset approach to seeding discussions. *Journal of Science Education and Technology, 18*, 321-333.

Clark, D., Sampson, V., Erkens, G. & Weinberger, A. (2007). Analytic frameworks for assessing dialogic argumentation in online learning environments. *Educational Psychology Review, 19*, 343-374.

*Choresh, C., Mevarech, Z. R. & Frank, M. (2009). Developing argumentation ability as a way to promote technological literacy. *International Journal of Educational Research, 48*, 225-234.

De Wever, B., Schellens, T., Valcke, M. & Van Keer, H. (2006). Content analysis schemes to analyze transcripts of online asynchronous discussion groups: a review. *Computers & Education, 46*, 6-28.

Hedges, L. V. & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego: Academic Press.

*Janssen, J., Erkens, G., Kirschner, P. A. & Kanselaar, G. (2010). Effects of representational guidance during computer-supported collaborative learning. *Instructional Science, 38*, 59-88.

*Joung, S. & Keller, J. M. (2004). *The effects of high-structure cooperative versus low-structure collaborative design of [sic!] decision change, critical thinking, and interaction pattern during online debates*. Paper presented at the Association for Educational Communications and Technology, 27th, Chicago, IL, October 19-23, 2004.

*Kollar, I., Fischer, F. & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning and Instruction, 17*(6), 708-721.

Munneke, L., van Amelsvoort, M. & Andriessen, J. (2003). The role of diagrams in collaborative argumentation-based learning. *International Journal of Educational Research, 39*, 113-131.

*Nussbaum, M. E., Sinatra, G. M. & Poliquin, A. (2008). Role of epistemic beliefs and scientific argumentation in science learning. *International Journal of Science Education, 30*(15), 1977-1999.

Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science, 328*, 463-466.

Raudenbush, S. W. (1994). Random effects models. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 301-321). New York: Russell Sage.

Schwarz, B. B., Neumann, Y., Gil, J. & Ilya, M. (2003). Construction of Collective and Individual Knowledge in Argumentative Activity. *The Journal of the Learning Sciences, 12*(2), 219-256.

*Seethaler, S. & Linn, M. (2004). Genetically modified food in perspective: an inquiry-based curriculum to help middle school students make sense of tradeoffs. *International Journal of Science Education, 26*(14), 1765-1785.

*Stegmann, K., Weinberger, A. & Fischer, F. (2007). Facilitating argumentative knowledge construction with computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning, 4*(2), 421-447.

*van Drie, J., van Boxtel, C., Jaspers, J. & Kanselaar, G. (2005a). Effects of representational guidance on domain specific reasoning in CSCL. *Computers in Human Behavior, 21*, 572-602.

*van Drie, J., van Boxtel, C., Erkens, G. & Kanselaar, G. (2005b). Using representational tools to support historical reasoning in computer-supported collaborative learning. *Technology, Pedagogy and Education, 14*(1), 25-41.

Walton, D., Reed, C. & Macagno, F. (2008). *Argumentation schemes*. Cambridge: Cambridge University Press.

*Weinberger, A., Stegmann, K. & Fischer, F. (2010). Learning to argue online: Scripted groups surpass individuals (unscripted groups do not). *Computers in Human Behavior, 26*, 506-515.

*Wiley, J. & Bailey, J. (2006). Effects of Collaboration and Argumentation on Learning From Web Pages. In A. M. O'Donnell, C. E. Hmelo-Silver & G. Erkens (Eds.), *Collaborative Learning, Reasoning, and Technology* (pp. 296-321). Mahwah, NJ: Erlbaum.

*Yeh, K. H. & She, H. C. (2010). On-line synchronous scientific argumentation learning: Nurturing students' argumentation ability and conceptual change in science context. *Computers & Education, 55*, 586-602.

Technology and Dialogic Space: Lessons from History and from the 'Argunaut' and 'Metafora' Projects

Rupert Wegerif, Yang Yang, University of Exeter, UK
Email: r.b.wegerif@exeter.ac.uk, y.yang@exeter.ac.uk

Abstract: In this theory paper we define dialogic space and outline its importance to computer supported collaborative learning. We argue that dialogic space is a complex concept combining transcendental and empirical aspects to account for the situated opening of opportunities for creative understanding in the tension between different perspectives. Despite this complexity dialogic space can be operationalized in concrete designs for learning. Some general principles are developed through a review of the literature on the indirect relationship between communications technologies and dialogic space. The EC funded Argunaut and Metafora projects to design online dialogic education environments are then used to illustrate more specific affordances of online design for a dialogic pedagogy.

Introduction

An account of dialogic education giving a central role to dialogic space as opposed to socio-cultural accounts that focus on mediation by cognitive tools was put forward by Wegerif in the first issue of the International Journal of Computer Supported Collaborative Learning (Wegerif, 2006). The concept of dialogic space was then elaborated in a book in the CSCL series, entitled Dialogic Education and Technology: Expanding the Space of Learning (Wegerif, 2007). The notion of dialogic space put forward in these two publications has proved of value for some empirical investigations, in particular several studies of Interactive Whiteboards (IWBs) in the UK have found the notion of dialogic space useful for theorizing the impact that this technology can have in classrooms (Mercer, Warwick, Kershner, Kleine Staarman, 2010: Scott, Ametller, Mortimer, and Emberton, 2010: Hennessey, S. In press). Others, however, have expressed concerns about the abstract nature of the concept asking how it can be operationalised and concerns that it does not give an adequate account of the active role that artefacts in supporting learning through educational dialogues (Hakkarainen, personal communication, Jerusalem, 2008). This paper is a response to these concerns. In it we outline exactly how being drawn into dialogic space is related to learning to think, how the concept of dialogic space can be operationalised in designs for CSCL and the affordances of designs with technology for supporting a pedagogy based upon the central notion of dialogic space. Although this paper is essentially a theory paper we also draw upon examples from two large funded projects in which designs for dialogic education with technology have been developed: the current Metafora project (www.metafora-project.org) funded by EC Framework 7 and the recently completed Argunaut project (www.argunaut.org) funded by EC Framework 6.

Dialogic Space and Education

Dialogic space is the space of possibilities that opens up when two or more incommensurate perspectives are held together in the creative tension of a dialogue. We use the term 'incommensurate' in this definition to emphasize the importance and the ineluctability of the dialogic gap. In real dialogues we speak from different embodied perspectives on the world and this difference cannot be simply resolved into unity because I cannot become you and you cannot become me. As Bakhtin points out on many occasions this difference is not a 'problem' but is constitutive of the flow of meaning in dialogues.

In analyzing why some groups were able to solve problems and other groups were not in primary schools it was argued that a key difference was the extent to which the way in which the children talked together opened up a shared space of reflection which allowed creative solutions to emerge (Wegerif, 2005). Because unsuccessful groups tended to identify with narrow images of a bounded self in opposition to others (disputational talk) or an image of the group itself as a harmonious entity (cumulative talk, from Mercer 2000) it was initially argued that the more successful talk involved an identification with the space of dialogue (Wegerif and Mercer, 1997). More recently it has been argued that this notion of identification with dialogue improving thinking is better understood in terms of a developing dialogue with the infinite other (Wegerif, 2010, p80). The idea is that while in a dialogue we might start just trying to persuade the other person, in doing so we inevitably listen to our own arguments as if from an outside point of view. This is the witness position, or the 'third' position, that Bakhtin writes is generated by every dialogue and refers to as the 'superaddressee' (Bakhtin, 1986). Bakhtin argues that words in dialogues are not just aimed at specific others they also become aimed at a sort of otherness in general. By following Bakhtin's logic the argument can be taken further towards a notion of dialogue with the infinite other or the idea of the outside. The argument is one of infinite regress. 1) Every dialogue generates a 'witness' position. 2) This witness position might appear to have a location but if you try to pin down this witness position in order to dialogue with it you will find that another witness position is automatically generated. 3) Because it can never be grasped, because it always runs away from us when we try to catch it, dialogue with the witness position leads on to dialogue with the infinite other. This analysis of infinity implicit in dialogues enables us to understand more clearly how children learn to reason. First they are called to explain themselves in dialogues with specific others. In the act of doing so, they become drawn into a dialogue with a third position that every dialogue generates, the position of the witness or super-addressee. While this might initially take a specific cultural form, such as the embodied norms of the community, every such fixed notion of the super-addressee can also be questioned leading to the idea of dialogue with the infinite other. The infinite other is not, of course, some kind of 'thing' but more like a constant call to go beyond prejudices and assumptions in order to see one's situation as if from the outside. Another way of putting this is that being drawn into dialogue with the infinite other is about acquiring a passion for truth (Wegerif, 2010, p81).

The significance of dialogic space for education can be understood through contrasting it to Vygotsky's notion of the Zone of Proximal Development or ZPD. With the ZPD Vygotsky introduced a particular version of the idea of a dialogic space into education. Through mutual attunement in the ZPD adults bring children to see things from their point of view thus leading children to reinterpret their initial spontaneous concepts through the more developed already existing 'scientific' concepts of the culture embodied in signs (Vygotsky, 1978). Vygotsky and his followers tend to present learning to think solely in terms to the internalization or personal appropriation of cultural tools such as signs through the medium of dialogue in the form of the ZPD. So for example Vygotsky claims that children learn to think logically through formal schooling, which leads them to personally internalize the categorizing practices of school science using cultural mediating means such as tables in text books etc. While dialogue in the ZPD is crucial for this vision of education this dialogue is assumed to take a rather limited and asymmetrical form. Vygotsky's account is a good account of how children learn bounded areas of thought such as, for example, how they may learn to use specific concept words appropriately or how they may learn to use an abacus to solve mathematics problems. However, it does not account well for how children might learn to think in an open-ended critical and creative way. To do this we simply need to augment the Vygotskian account of learning with the understanding that children do not only internalize or personally appropriate cultural tools such as logical tables and concept words, they also internalize or appropriate the dialogic space through which such signs are able to mean in the first place. Learning to think for oneself is learning to carry around a space of dialogue through which anything and everything can be questioned and seen in a new way.

So a concept of dialogic space is essential for understanding education into thinking and creativity but what exactly is dialogic space? Clearly it is not visible or tangible in a direct empirical manner. Looking from the outside we may see people apparently in dialogue but we cannot directly see the dialogic space of possible meanings that opens up within the dialogue. From the outside we can say that dialogues are situated in space and time, in culture and in history, but from the inside they escape situation. The virtual landscape of the inside of the dialogue might well refer to the past and the future as well to places at a distance. The meaning of the cultural and historical context we construct in order to situate dialogues is always itself a matter for dispute within dialogues. If, as Bakhtin argues, meaning is only possible in the context of a dialogic gap then it follows that dialogic space is a transcendental concept. Transcendental, as Kant defined the term, means a pre-condition for experience rather than an object encountered within experience. We all have experience of meanings in dialogues and those experiences imply a prior opening of dialogic space that makes them possible. The notion of the infinite other outlined earlier suggests that, contrary to sociocultural orthodoxy, there is an important aspect of all dialogues that is universal, unsituated and atemporal. Viewed from the outside all dialogues are different but experienced from the inside they all share something in common which is the infinite potential to be drawn into self-questioning and reflection which we referred to as the idea that the infinite other is a potentially emerging voice within all dialogues. Bakhtin developed the notion of 'Great Time', a time and space in which all voices can communicate with each other, to explain how it was that he, for example, a 20th century Russian, could enter into constructive dialogue with, for example, the voice of Socrates, a Greek from the 5th Century BCE. However great the differences in culture and history it seems that creative dialogic engagement is always possible and this is so because there is something infinite at the heart of dialogue.

If dialogic space has a transcendental aspect it also has an empirical aspect. Classroom research described in Wegerif (2007) shows that dialogic spaces can be opened up where they did not exist before through the simple procedure of interrupting children's engagement with an interactive game in order to ask them open questions that lead them to predict and reflect. This opening of dialogic space is not directly visible but it is easy to deduce (or, more strictly speaking, to abduce) from the visible indicators such as transcripts of talk. Merleau-Ponty (1968) uses the phrase 'the invisibles of this world' for forms like dialogic space that cannot be directly seen but are not unknowable transcendentals because they have a direct effect on the visible world such that they can easily be seen behind it. Dialogic space then is quasi-transcendental, it is transcendental but also empirical in that it can be opened, widened and deepened in ways that are indirectly measurable empirically. One way to conceptualise this is that dialogic space is an opening within a highly structured surface of an underlying space of possibilities. Depending on the context that opening could be more or less limited. All dialogues increase the degrees of freedom of thought to some extent, all also have the infinite potential for new meaning described above through the mechanism of the infinite other, however, in practice most dialogues remain relatively bounded to a limited range of alternatives within a clear cultural context.

A Brief History of Communications Technology and Dialogic Space

There is an interesting strand of literature on the intellectual affordances of communications media which can help us understand the relationship between technology, dialogic space and education. In the following very brief review of this literature we will look at what has been claimed for the impact of oral dialogue, written media and the internet on thinking in order to illuminate how dialogic space is shaped by technology and to approach a preliminary understanding of some of the possible pedagogical affordances of new communications media.

Socrates notes in the Phaedrus that the living word of face-to-face dialogues has the potential for stimulating understanding in others in a chain that is endless and so produced 'immortality'. He contrasts this living word of dialogue to the dead words of writing that are just like shadows or ghosts because they are not inhabited and cannot answer back. Many have noted that face to face dialogue assumes a certain 'mutual attunement' between participants (e.g Rommetveit, 1992). Utterances in dialogues do not stand alone but they respond to previous utterances and they are designed to influence the person addressed. In a dialogue, in other words, the other (understood here as the addressee), is not simply outside me but appears on the inside of me shaping my utterances from within even as they form. Even to engage in dialogue I need to be able to see myself to some extent from the other's point of view. Socrate's account of the intellectual power of the living word remains valid today and is a rebuke to those who claim that our thinking is essentially mediated by our technology. If this were true then our thinking would be obviously superior to that of the pre-literate Socrates (as reported to us by his literate student Plato) but in fact this is not the case. However, Socrates shows a certain naivety as to the impact and limitations of the spoken word as a medium of dialogue. The limitations of oracy as a medium of thought have been brought out by others in more established literate cultures.

In oral cultures words are only found in the ephemeral context of face-to-face speech. By the time I have grasped the import of my interlocutor's words they have vanished and I cannot turn back to re-examine them. This means that words, and the ideas they carry, are inevitably closely bound up with time and place. Some have argued, for example, that without literacy there are no universal abstract concepts (e.g Ong, 1982; Olson, 1994: Goody, 1977). Writing, it is argued, enabled ways of thinking that were not possible with face-to-face dialogue alone. One example is the way in which the 'religions of the book' could disembed themselves from a physical context to cross seas and mountains and claim adherents in different cultures. They could separate their truth from places because it was contained in words and so it became transportable (Goody, 1977). The Christian Bible and the Koran are known as 'the word of God'. The spread of Buddhism also depended in the writing down of the Pali Canon and its transportation over mountains and seas to new lands.

As we noted above, Socrates, an oral thinker, is reported as criticizing writing precisely for taking the idea of truth away from the living context of words in face to face dialogues and claiming truth for what he referred to as the dead words of writing (Plato, 2006). In the Christian Holy Bible there is an interesting passage that appears to state that the writing is now closed and anyone adding a word to it will be cursed (The Bible: Revelation 22:18-9). If so then this is indicative of a new idea of truth that arrives with writing. Truth here is perhaps being seen as a closed finished thing of universal relevance separate from a context of utterance.

Of course literacy did not suddenly take over nor did it ever completely replace oracy, but at a certain point, according to Toulmin, it seems to have become the dominant medium for our self-understanding of the nature of thought. Toulmin investigates the nature of 'modernity' and finds it in a shift from respecting dialogic modes of thought to respecting only written modes of thought. Before 1600, he writes, both rhetoric and logic were seen as legitimate modes of philosophy (Toulmin, 1992, p30). He contrasts Montaigne's highly contextualised and dialogic brand of philosophy to the abstract universal certainty sought after by Descartes. After Descartes there was a shift from seeing truth in terms of utterances in dialogues in situations to seeing truth in terms of propositions and proofs that were unsituated and universal (1992, p31). In other words modernity can be characterised by 'monologicality', the assumption that there is only one true perspective or voice, which in turn is an effect of the dominance of written modes of reasoning over oral modes.

Harold Innis, the Toronto based communication theorist who inspired Marshall McLuhan, draws attention to the requirement of empires to have portable written communications. He locates the development of the technology of writing the struggle of empires to impose a uniform, written, law. Writing enabled empires since the very idea of empire is to be able to write the law code at the centre and spread the same law uniformly out to all the provinces (Innis, 1950, p30). The first written law code, that of Hammurabi (2123-2083 BC), served the purpose of centralising power. Innes's detailed accounts show how communications technologies such as paper and print, even horses and ships carrying papyrus scrolls, were all essential to empires and shaped the nature of those empires.

Television and Radio, although electronic media, shared the same one to many nature of print media and so were easily seen by Adorno and Horkheimer (Horkheimer and Adorno, 1972) as continuing the imperialistic tendencies of print. The internet, however, is very different, especially the so-called Web 2.0 iteration of the internet. The internet facilitates participation and the same kind of two-way or multi-way dialogue found in face to face talk but it also supports the communication over distance found in writing.

If orality and literacy impacted on ways of understanding thinking then what impact will the internet have? It is too early to say. While oracy and literacy have had millennia to shape collective cognition, the widespread use of the internet is still just beginning. One possible impact noted by Gabi Salomon some years back (1992) as the 'butterfly effect' and recently made into a best-seller by Nicholas Carr, 'The Shallows: What the Internet is doing to our brains' (Carr, 2010), is to make us all more superficial and distracted. Whereas reading books takes commitment and can lead to depth understanding, use of the internet encourages browsing nuggets of pre-processed information condemning us to superficiality, or so the argument goes. This analysis fits reasonably well with those who argue that the rise of the internet marks the end of the modern self, said to be individual and autonomous. Mark Poster, for example, argues that:

> Electronic culture promotes the individual as an unstable identity, as a continuous process of multiple identity formation and raises the question of a social form beyond the modern, the possibility of a post-modern society (1995, p398)

The claim that new communications technology will usher in a post-modern or post-structuralist reality of fragmented, and multiple identities sounds a bit negative but there is a more positive sounding corollary. The other side of the break-down of the authority of the author is individuals becoming more open to dialogue with others and with otherness. If we accept Toulmin's account that a focus on print has had a monologic effect, turning 'utterances' in dialogues into 'propositions' in proofs; is seems possible then that the internet can restore us to a more participatory and dialogic way of understanding thinking. However, while the internet supports dialogue this is different from the oracy that preceded literacy, for one thing this is no longer a dialogue limited to a physically located community but a dialogue without any necessary spatial limits. David Barton tells an interesting story of how the way he saw the hills near his home was subtly altered after he received comments on his photographs of those hills from correspondents in Germany and Japan via the Web 2.0 photograph sharing site, Flickr (see Lee and Barton, 2011). Ong has argued that practices around writing and reading books led to the formation of a sense of an individual inner autonomous space that contrasted to the more collective identity of selves in oral societies and enabled critical thinking leading to political change. It seems plausible that some forms of blogging promote similar kinds of 'inner space' capable of standing back from and criticizing tradition, but in a collective form without the sense of individual autonomy.

One clear lesson that can be learnt from the literature about the impact of modes of communication on thinking and society is that mentality is not just a causal effect of the technology. Ong brings out how one way of writing and reading can cement communal solidarity, the reading aloud of a manuscript such as the bible which was common in the middle ages (Ong, 1982, p117), whilst another way of writing and reading, silent and solitary writing and reading of books, can support the formation of a separate autonomous inner self able to stand back from the culture around it (Ong, 1982, p129). The message we can take from this is that the apparent fragmentation and superficiality induced by internet use according to Carr and others is not an inevitable effect of the internet but a possible consequence of one way of using this new technology. Just as the previously dominant media of communication, oracy and literacy, can be a part of cultural practices that have quite different effects on thinking, so then can the internet. This analysis suggests a possible role for educational research as determining what are the pedagogical affordances and dangers of the internet and how it can best be integrated into the practice of education.

Argunaut and Design for Dialogic Space

The history of communications technology gives us some clues as to how different technologies afford different possibilities of thought. Writing things down may have led to the delusion of unsituated and ahistorical truth but it also enabled deeper reflection because ideas perdured over time and so could be returned to as artefacts to be re-animated by new perspectives and collectively developed. In other words writing, in the form of journals for example, supported a wider and deeper dialogue on scientific issues than oracy alone could manage. However an aura of quasi-religious closed truth still hangs around printed scientific texts making drawing students into the practice of science as dialogue difficult.

New communication technologies and tools offer many new affordances for dialogue. Computer mediated dialogues expand the 'space' of dialogue by spatialising time so that many can 'talk' in parallel and their different voices can be represented by spatial differences in an interface. Normally this different way of doing dialogue is represented in a kind of traditional playscript with one utterance after another listed in a temporal sequence. This linear list is a kind of metaphor for the progression of moments in time. Even this arrangement however makes it easy to lose the context of the argument. The Knowledge Forum is an early example of an interface that shifts the dialogue representation from this linear form, one utterance after another, form to a more visual arrangement on a plane more like a concept map. The same move is made by Digalo, an online dialogue environment developed by the EC funded Dunes project and used as part of the Argunaut project) relies on argument maps.

The Argunaut system developed during the project uses the graphical e-discussion environments Digalo (dito.ais.fraunhofer.de/digalo/) and FreeStyler (www.collide.info/software) for students, along with a Moderators Interface (MI) for teachers, which includes a range of awareness indicators and tools for intervention designed to make the task of moderation easier. As well as providing awareness of relative participation, types of messages and the relationships between people through social network diagrams, we also sought to provide awareness indicators for the *quality* of the discussion. Using this system we were able to investigate the hypothesis that the spatialized reasoning of dynamic concept maps supported creative reasoning. (Wegerif, McLaren et al, 2010) We investigated the impact dynamic concept mapping on creativity using a coding for creativity based on a pattern-matching algorithm combined with stimulated recall interviews of participants. Through the interviews we found that the non-linear nature of the maps with multiple ideas co-present stimulated creativity. Our findings were that: 1) the emergence of new perspectives in a graphical dialogue map can be coded for using an intuitive interpretation of the construction of meaning in a dialogue in a way which can be recognized reliably by a machine algorithm; 2) that the trigger events leading to the emergence of new perspectives are most commonly open questions and disagreements and 3) that the spatial representation of messages in a graphically mediated synchronous dialogue offers a pedagogical affordance for creativity. This research supports the view that some normative ideas of argument, even those informed by the idea of dialogue such as the dialectic progression from thesis and antithesis towards a synthesis, have been shaped by the linear text form of writing that is always moving towards the conclusion at the bottom of the page. Dynamic multi-user concept mapping on the other hand also supports thinking but has a greater affordance for the creative emergence of unanticipated new ideas that comes from the co-presence of multiple perspectives without any possibility of overcoming one by another or any sense of the need to move together towards the bottom-line.

Metafora and the Use of Tools to Represent Voices

The follow-up Metafora project is investigating the combination of using induction into online mediated dialogic space as a way of teaching general thinking and creativity. The three year project, funded by the European Commission, is developing new theories and pedagogies to increase learners' awareness of learning together, where children aged 12-16 explore authentic science challenges with classmates. The challenges are based around science topic themes, which are key issues of 21st century science curriculum, requiring children to develop general thinking strategies of new knowledge age. To exploit collaboration and discussion, the challenges will be ill-defined and allow different points of views and priorities.

Metafora promotes dialogic learning in three domains: 1) learning to learn together (L2L2); 2) learning to argue and 3) learning about specific mathematics and science concepts. For the learning to learn together component we have developed and trialed an iconic language standing for different activities and emotional attitudes to help the students, working in groups of 5 to 10, to plan their learning. Using a planning space, students iteratively model their own learning activities and reflect upon them. We are investigating if working with these planning and reflection tools in the planning space raises awareness of the stages and attitudes implicit in successful group shared enquiry over time thus help scaffold novices in L2L2 towards becoming experts.

The model of learning to learn together (L2L2) requires that students become aware of the key variables in successful learning together and learn how to use them. This is like learning to recognize and use a new language. From the literature we propose that some of the key variables include staging or sequencing activities and different types of dialogue. But for young children to learn effective dialogue it is necessary for them to become more aware of the way in which they talk or interact which includes becoming aware of and consciously regulating their emotional attitudes to each other and to the topic.

From a dialogic perspective cognition cannot be divorced from affect because cognition is an emergent property of dialogues, and dialogues are defined not by the exchange of signs, something which could be quite emotionally neutral, but by the intersubjective space of relationship. Relationships always have emotional tone. One of the main reason children and young people fail to learn to learn together in groups is because they fail to recognize and take responsibility for the impacts of their emotions and attitudes towards each other and towards the task. In designing icons to help them become aware of and take responsibility for their attitudes we realized that we were using tools in an interesting way. Some socio-cultural literature has suggested that cultural voices become appropriated by individuals to be used as cognitive tools. In Metafora we are designing tools that evoke voices. Our icons are representations of attitudes such as being critical, playful, visionary, ethical and so on, which the users have to inhabit in order to speak with the voice of that attitude.

Discussion: the Pedagogical Affordances of CSCL for Dialogic Education

Dialogic education combines teaching through dialogue with teaching for dialogue. As well as learning specific facts and specific skills, dialogic education draws learners into dialogue. Dialogic education at both the individual and collective level seeks to draw learners from what Bakhtin called 'narrow time', a concern only with the local and the short-term, to what he called 'Great Time', the dialogue of all times and all cultures. In classrooms this can be seen in moving children from the structured surface of physical time and space in which things happen but without reflection or the possibility of change towards the opening, widening, deepening and heightening of dialogic spaces. A key mechanism for heightening the dialogue is that of taking the perspective of the superaddressee or witness in order to reflect on what has been said or claimed. This mechanism within dialogues can shift participants from trying to persuade specific others to trying to understand in general.

If dialogic space is the main medium of education and learning to think then what is the role of technology? The answer is that communications technology opens and shapes different kinds of dialogic spaces with different affordances for thought and for education. Our brief accounts of oracy and literacy offered arguments that the medium of communication does make a difference to the possibilities of thought and to the direction of education. It is as if the medium itself becomes a key voice in the dialogue. These accounts also suggested that while technology does offer a framework of constraints and affordances its actual impact on thought and on education will depend on how it is used or how it is embedded in cultural practices. This point is particularly important in understanding the potential impact of the internet age as the internet is not one technology but many bundled up together offering a great variety of different possible uses. The potential dangers of the internet for the quality of thought have been pointed out but reflection on the nature of the internet shows that it also had affordances to support a dialogic education, opening new spaces for dialogue, widening them by introducing diverse voices, deepening dialogue through the provision of continuity in the form of malleable artifacts that allow reflection and analysis and also heightening the dialogue in the direction of dialogue with the infinite other through the reflection that occurs in dialogue across difference.

Design-based research in developing the Argunaut and Metafora systems illustrates how designs for technology can shape dialogic space in a way that has an impact on the type and quality of thinking and so on the possibilities of education for thinking. The research described with Argunaut demonstrated the simple point that the way in which we shape the space within which people communicate has an impact on their thinking. In this case dynamic concept-mapping was shown to support the creative emergence of new ideas through the juxtaposition of multiple perspectives on a single topic. The provisional work-in-progress report from Metafora suggested the significance of designing for voices and for the emotional colors of dialogic space. This represents a shift in pedagogical design principle from using voices as tools for thinking to using tools as voices for thinking.

In conclusion then we are arguing that:

- dialogic space is an opening of reflection and possible new perspectives around a dialogic gap;
- dialogic analysis helps us to understand education for thinking as shifting orientations and drawing learners from dialogue with specific others towards dialogue with the infinite other;
- communications media have different constraints and affordances in relation to dialogic space;
- the design of communications media and spaces offers affordances for different kinds of thinking and so for different kinds of education into thinking;
- artefacts can act like voices in a dialogue supporting continuity, reflection and cumulative development;
- analysis only in terms of tools and 'construction' can miss the point that education, amongst other things, is about helping new participants in dialogue acquire their own unique voice
- in a project to teach learning how to learn together we are exploring what happens when we embody key dialogic orientations in icons which then serve as tools to support the acquiring of voices

References

Bakhtin, M. (1981). Discourse in the Novel. In M. M. Bakhtin. The dialogic Imagination. Four essays by *M. M. Bakhtin*. Austin: University of Texas Press, pp. 259-422.

Bakhtin, M. (1986) Speech Genres and Other Late Essays. Austin:University of Texas.

Carr, N. (2010) The *Shallows*: What the Internet is Doing to Our Brains; New York, NY: W. W. Norton & Co.,

Goody, J. (1977) The domestication of the savage mind. Cambridge, England: Cambridge University Press.

Hennessey, S. (in press) The role of digital artefacts on the interactive whiteboard in supporting classroom dialogue. Journal of Computer Assisted Learning.

Horkheimer, M and Adorno. T (1972). *Dialectic of Enlightenment*. New York: Seabury.

Innis, H. (1950) *Empire and Communications*. Oxford: Clarendon Press
Lee, C. & D. Barton (2011) Constructing Glocal Identities through Multilingual Writing Practices on Flickr.com. *International Multilingualism Research Journal,* 5(1), 39-59.
Mercer, N, Warwick, P, Kershner, R, Kleine Staarman, J (2010) Can the interactive whiteboard help to provide 'dialogic space' for children's collaborative activity?. Language and Education ()
Mercer, N. (2000). *Words and Minds: How We Use Language to Think Together*. London: Routledge.
Merleau-Ponty, M. (1968). The Visible and the Invisible (Edited by Claude Lefort and
Olson, D. (1994) *The World on Paper*. Cambridge: Cambridge University Press.
Ong, W. J., 1982, *Orality and Literacy: The Technologizing of The Word*. Methuen, London.
Plato, (2006) *The Phaedrus*. Translated by Benjamin Jowett. Online. Available http://ebooks.adelaide.edu.au/p/plato/p71phs/phaedrus.html (accessed 1st August 2008)
Poster, M. (1995) *The Second Media Age* Oxford: Blackwell
Rommetveit, R. (1992). Outlines of a dialogically based social-cognitive approach to human cognition and communication. In A. Wold (Ed.). *The dialogical alternative: towards a theory of language and mind* (pp. 19- 45). Oslo: Scandanavian Press.
Salomon, G. (1992). New information technologies in education. In M. C. Alkin (Ed.), *Encyclopedia of educational research (6th Ed.)* (pp. 892-903). New York: Macmillan
Scott, P., Ametller, J., Mortimer, E., & Emberton, J. (2010). Teaching and learning disciplinary knowledge: Developing the dialogic space for an answer when there isn't even a question. Chapter 14. In K. Littleton & C. Howe (Eds.), *Educational Dialogues: Understanding and Promoting Productive interaction*. Abingdon, Oxon: Routledge.
Toulmin, S. (1990) *Cosmopolis: The Hidden Agenda of Modernity*. New York: Free Press.
translated by Alphonso Lingis). Evanston, Il: Northwestern University Press.
Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.
Wegerif, R (2007) *Dialogic, Education and Technology: Expanding the Space of Learning.* NY and Berlin: Springer.
Wegerif, R (2010) Mindexpanding: teaching for thinking and creativity in primary education. Buckingham: Open University Press.
Wegerif, R. (2005) Reason and creativity in classroom dialogues.*Language and Education* 19(3), 223-238
Wegerif, R. (2006). A dialogic understanding of the relationship between CSCL and teaching thinking skills. International Journal of Computer Supported Collaborative Learning 1(1), 143-157.
Wegerif, R., McLaren, B. M., Chamrada, M., Scheuer, O., Mansour, N., Mikšátko, J., et al. (2010). Exploring creative thinking in graphically mediated synchronous dialogues. *Computers & Education, 54*(3), 613-621.

The Relationships among Online Question-Generation, Peer-Assessment and Academic Achievement

Fu-Yun Yu, National Cheng Kung University, Taiwan, fuyun.ncku@gmail.com
Chun-Ping Wu*, TamKang University, Taiwan, cpwu303@gmail.com

Abstract: This study examined the relationships among student performances in question-generation, peer-assessment, and academic achievement in four different online peer-assessment learning environments. Eight fifth-grade classes (N=253) were randomly assigned to one of four different identity revelation modes (real-name, anonymous, nickname, and self-choice) and participated in the study for six weeks. An online learning system that allows students to contribute to and benefit from the process of question-generation and peer-assessment was adopted. Data analysis revealed significant relationships among the three examined variables. Additionally, the identity revelation modes moderated the predictive power of online question-generation performance on academic achievement. Specifically, the predictive power in the self-choice mode was the greatest. Finally, self-choice identity mode strengthened the relationships between peer-assessment and question-generation performances. The empirical significance of the study and suggestions for learning system development and instructional implementation are provided.

Introduction

In response to the call for the diversification of question sources and an emphasis on engaging students in the process of learning and knowledge construction, student question-generation has been increasing in popularity over the past decades (Brown & Walter, 2005; English, 1998; Silver, 1994). Studies on student question-generation have substantiated its efficacy in enhancing learners' retention of information, reading comprehension, motivation, in-group communication and interaction, satisfaction with past learning experiences, and problem-solving abilities (Abramovich & Cho, 2006; Barlow & Cates, 2006; Brown & Walter, 2005; Dori & Herscovitz, 1999; Leung & Wu, 1999; Silver, 1994; Yu & Liu, 2005; Yu, Liu, & Chan, 2005). Even though most studies have supported the student question-generation approach, investigations into factors that may affect learning performances during student question-generation activities are scare.

The effects of differences in personal characteristics and technology integration methods on adoption and use of student question-generation have been examined and yielded useful suggestions for instructional implementation (Wu & Yu, 2009; Wu & Yu, 2010). In view of the fact that students normally have limited experience generating questions during formal schooling (Moses, Bjork, & Goldenberg, 1993; Vreman-de Olde & de Jong, 2004; Yu & Liu, 2009), their performance in the task itself (i.e., question-generation) and its relationship to academic achievement in the applied contents is one area worth examining.

Moreover, to allow question-authors to receive timely and personalized feedback about contributed question items, peer-assessment is coupled with question-generation activities in most systems (Yu, 2009). Empirical evidence, which spans more than two decades, supports its facilitative effects for learners' motivation, higher-order thinking, cognitive re-structuring, level of performance and attitudes (Brindley & Scoffield, 1998; Falchikov & Goldfinch, 2000; Gatfield, 1999; Hanrahan & Isaacs, 2001; Purchase, 2000; Topping, 1998; Tsai, Lin, & Yuan, 2002; van Gennip, Segers, & Tillema, 2009; Venables & Summit, 2003; Wen & Tsai, 2006; Wen, Tsai, & Chang, 2006). However, students' performances in peer-assessment activities itself in relation to their learning (as reflected in their academic achievement) have rarely been investigated.

In an attempt to gain some knowledge on these areas, a study was conducted to examine whether students' performances in question-generation are related to academic achievement. Moreover, issues on what the relationships between peer-assessment performances and academic achievement and whether students with better question-generation performances offer peers better comments were also examined. Finally, since different forms of user identification in student question-generation and peer-assessment has been reported to lead to different psychological reactions (Yu & Liu, 2009), whether user identification moderates the relationships among question-generation, peer-assessment and academic achievement was investigated. The research questions of this study are:

1. Are learners' question-generation performances correlated with academic achievement? If the hypothesized correlation is supported, will the relationship vary with the use of different identity revelation modes?
2. Are learners' peer-assessment performances correlated with academic achievement? If the hypothesized correlation is supported, will the relationship vary with the use of different identity revelation modes?
3. Are learners' question-generation performances correlated with their peer-assessment performances?

Method

Two hundred and fifty-three 5th graders from eight intact classes participated in the study for six consecutive weeks. Students were informed that the introduced activity was intended to support their science education. To ensure that participants possessed the fundamental skills for the activity, a training session about generating questions and peer-assessment, including hands-on activities, was held at the beginning of the study. For the duration of the study, students were directed to individually generate at least one question in accordance with the instructional content covered and assess at least two questions each week after attending three instructional sessions allocated for science.

A learning environment that allows students to contribute to and benefit from the process of constructing question items, as well as exchanging ideas with their peers about the composed questions, was used. The system allowed interacting parties to communicate back and forth electronically about a specific question item. Essentially, assessors gave evaluative feedback using an online assessment form (with assessment criteria and type-in open space) (see Fig. 1). Alternatively, question composers were able to provide an elaborated explanation for their assessors to further respond to (the very moment the feedback form was submitted) (see Fig. 2). In short, the system allowed both of the interacting parties to engage in argumentative dialogue.

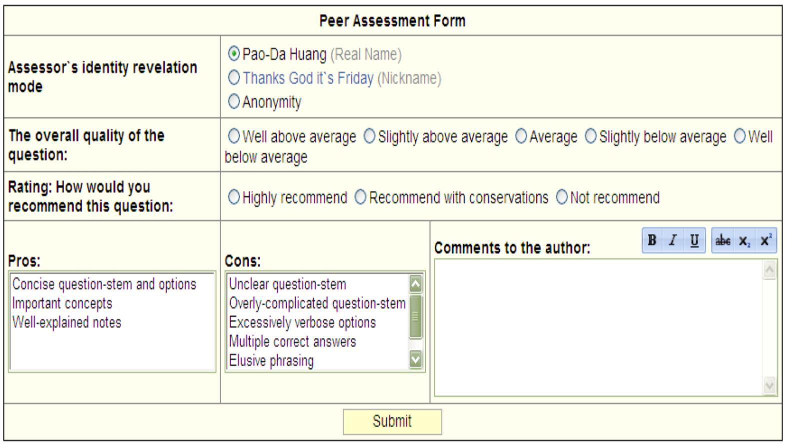


Figure 1. Assessment Form for Assessors to Provide Feedback to Question-Authors.

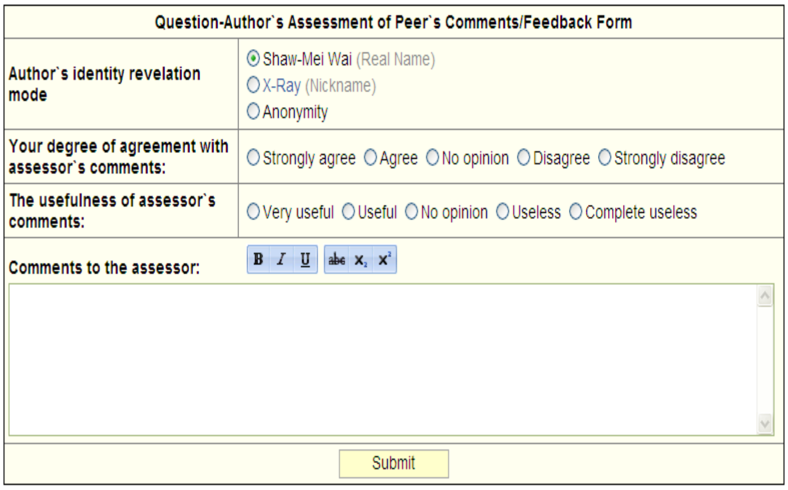


Figure 2. Assessment Form for Question-Authors to Communicate with Assessors.

Students' performances in question-generation and peer-assessment were defined as all questions students generated and assessed during the study. To assess students' performances in each respective activity, a set of criteria were developed. Specifically, in reference to the Torrance creativity index, King's question cognitive levels (1992) and questions generated by students, the following criteria were adopted: fluency, flexibility, elaboration, originality, cognitive level and importance. Each of the indexes was further operationally defined to ensure objective assessment. For example, the fluency index (score ranging from 0 to 3) was determined by correctness, clarity and conciseness of the composed question. A composite score was created for each generated question. As for peer-assessment, it was evaluated in terms of 3 discrete levels: (1) quality judgment, (2) quality judgment with identification of strong and weak areas, and (3) quality judgment with explicit suggestions for further refinement of questions. Students' academic performance was based on the mid-term and final exams, which were centrally administrated at the participating school.

To test the moderation effects of the different identity revelation modes, four treatment conditions were devised. In the real-name condition, the student's full-name was retrieved automatically from the database and shown on screen when questions and feedback were viewed (See Figure 3). In the anonymous group, information about the question-author and assessor were not shown, and only the word "anonymous" was marked (See Figure 4). In the nickname group, the student's self-created identity was shown at the top of the generated questions and rendered comments (See Figure 5). Finally, to take into account the recent finding that participants exhibited statistically significant different preferences for different user identity revelation modes when rendering comments (Yu & Liu, 2009), a self-choice group was created. In the self-choice group, rather than being assigned a specific and fixed identity revelation mode, the identity mode used was chosen by the user (see Figs. 1 & 2 top portion: identity revelation mode choice space). Students were free to change to whichever mode each time they generate or assess a new item. Eight participating classes were randomly assigned to four different treatment conditions (i.e., two classes per group).

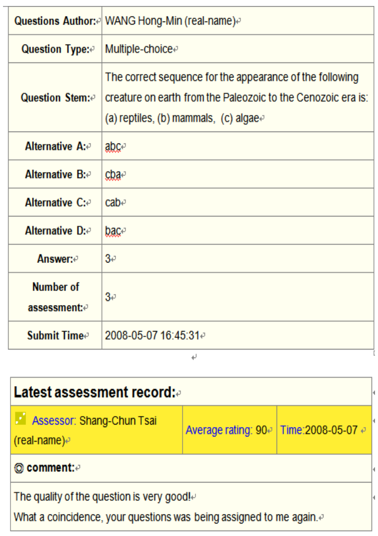


Figure 3. Real-Name Mode.

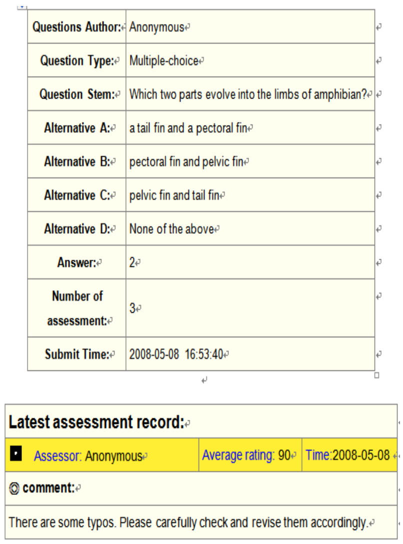


Figure 4. Anonymity Mode.

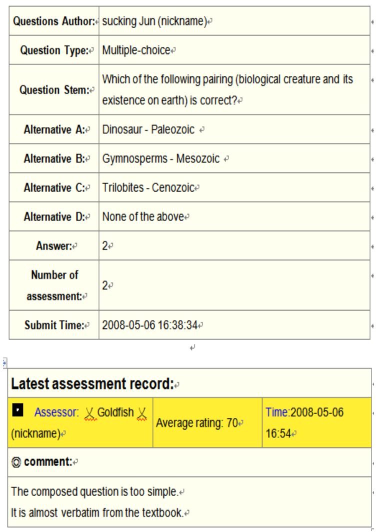


Figure 5. Nickname Mode.

Results

The means and standard deviations of students' performances in generating questions, peer-assessment, and midterm and final exams are listed in Table 1. The potential effects of gender and prior experience in the introduced strategies have been reported (Topping, 2010). As can be seen in Table 1, gender and prior experience in either online question-generation or online peer-assessment were approximately the same across different treatment conditions. Data was analyzed using partial correlation while controlling for the influence of the mid-term exam scores.

Table 1: Descriptive statistics of the examined variable.

Treatment	Gender		Online QG Experience		Online PA Experience		QG Performance	QA Performance	Mid-term	Final exam
	Male	Female	Yes	No	Yes	No	Averaged Mean(SD)	Averaged Mean(SD)	Mean (SD)	Mean (SD)
Real- name (n=65)	33	32	25	40	26	39	5.02 (2.37)	4.50 (2.91)	85.48 (7.74)	91.32 (6.75)
Anonymous (n=63)	32	31	35	27	35	26	6.07 (2.48)	5.23 (3.20)	86.10 (8.83)	92.24 (7.08)
Nickname (n=64)	32	32	25	39	27	37	4.52 (2.25)	3.96 (2.74)	84.41 (9.6)	91.75 (8.29)
Self-choice (n=61)	28	33	29	32	27	34	4.54 (2.39)	4.15 (2.65)	87.39 (9.86)	90.57 (10.07)
Total (n=253)	125	128	114	139	115	138	5.04 (2.44)	4.46 (2.91)	85.82 (9.04)	91.48 (8.10)

Note: QG refers to Question-Generation; PA refers to Peer-Assessment

Relationships between Question-generation Performances and Academic Achievement

The partial correlation analysis results indicated a significant relationship between performances in generating questions and academic achievement (r= 0.14, p = 0.02). A hierarchical regression analysis was conducted to explore the moderation effect of different identity revelation modes on the strength of the relationship. The analysis showed that performances in question-generation and the different identity revelation modes interacted significantly with regard to their relationships with achievement (β = 0.34, p =0.03). As shown in Table2, the significant effects of the mid-term exam, the treatment and the interaction between question-generation performances and treatments on achievement were found. However, a direct effect of question-generation performances on achievement was not found. In short, identity revelation modes fully moderated the predictive power of question-generation performances on achievement. A series of follow-up regression tests, which uncovered the moderation phenomena, indicated that the question-generation performances of students in the self-choice mode significantly positively predicted achievement scores (β_{opt} = 0.20, p= 0.04) and that the predictive power of the self-choice group was the greatest of the four modes.

Table 2: Hierarchical regression model summary of the question-generation performance.

Model Variable	Achievement (Dependent Variable)		
	Model 1	Model 2	Model 3
Controlled Variable Mid-term Exam	0.56**	0.56**	0.56**
Independent Variable Question-generation performance Treatment	0.12*	0.11* -0.05	-0.14 -0.28**
Interaction Treatment X Question-generation performance			0.34*
R-square F Sig	0.36 70.99 0.00	0.36 47.60 0.00	0.36 37.56 0.00
ΔR-square ΔF Sig	0.36 70.99 0.00	0.00 0.88 0.35	0.01* 5.09 0.03
Note: a. predictor:(constant), Mid-term Exam (controlled), Question-generation performance b. predictor:(constant), Mid-term Exam (controlled), Question-generation performance, Treatment c. predictor:(constant), Mid-term Exam (controlled), Question-generation performance, Treatment, interaction			

Relationships between Peer-assessment Performances and Academic Achievement

The partial correlation between performances in peer-assessment and achievement is statistically significant (r= 0.15, p = 0.02). However, the moderation effects of the different identity revelation modes on the strength of the relationship was not supported by the hierarchical regression analyses, as reported in Table3 (β = 0.1, p =0.48).

Table 3: Hierarchical regression model summary of peer-assessment performance.

Model Variable	Achievement (Dependent Variable)		
	Model 1[a]	Model 2[b]	Model 3[c]
Controlled Variable Mid-term Exam	0.56**	0.56**	0.56*
Independent Variable Peer-assessment performance Treatment	0.13**	0.12* -0.05	0.04 -0.11

Interaction Treatment X Peer-assessment performance			0.1
R-square	0.36	0.37	0.37
F	71.44	48.00	36.05
Sig	0.00	0.00	0.00
ΔR-square	0.36	0.00	0.001
ΔF	71.44	1.07	0.49
Sig	0.00	0.30	0.48
Note: a. predictor:(constant), Mid-term Exam(controlled), Peer-assessment performance b. predictor:(constant), Mid-term Exam(controlled), Peer-assessment performance, Treatment c. predictor:(constant), Mid-term Exam(controlled), Peer-assessment performance, Treatment, interaction			

Relationships between Question-generation and Peer-assessment

The correlation analysis result showed a significant relationship between performances in generating questions and peer assessment (r= 0.66, p <0.01). Additionally, real-name, anonymous, nick-name, and self-choice all reached statistical significance with scores of r= 0.69 (p <0.01), r= 0.58 (p <0.0), r= 0.61 (p <0.01) and r= 0.71 (p <0.01), respectively. The relationship with the self-choice mode was the strongest.

Conclusion and Implications

In this study, relationships among question-generation, peer-assessment and academic performances were first explored, followed by an examination of how different identity revelation modes strengthen or weaken the relationships. The current study confirmed that the correlations among the three examined variables were significant. First, this study substantiated that students who performed better at question-generation tended to perform better in academic achievement. Additionally, the results, which supported the moderation effects of the identity revelation modes, evidenced that students in the self-choice mode who performed better in question-generation could be expected to perform better in academic achievement. Second, correlations between students' performances in peer-assessment and academic performance were also indicated for all four different identity revelation modes; however, the intensity of the relationships was similar in different modes. Third, the study confirmed that there was a high degree of correlation between question-generation and peer-assessment, and the intensity of the relationship was strongest in the self-choice mode.

These findings have important empirical significance as well as implications for online system developments and instructional implementation. First, the current study substantiated the relationships between performances in question-generation, peer-assessment and academic achievement on the applied content. In light of these findings, instructors interested in using the student question-generation approach for promoting academic performance are advised to address the issue of supporting question-generation and peer-assessment performances by, for instance, including deliberate training or building question-generation or peer-assessment scaffolds (Yu, 2009; Rosenshine, Meister, & Chapman, 1996), etc.

Furthermore, moving away from assessee's subjective perceptions of gains and attitudes towards peer-assessment, this study included objective measures (learning gains in academic contents) and evaluate its relationship with students' performances in question-generation and peer-assessment. In this study, students with better question-generation performances tended to offer peers better feedback and excelled better in academic assessment afterwards. A word of caution, however, is warranted. Since the activities of question-generation and peer-assessment were implemented simultaneously, no cause-and-effect but correlations can be inferred from the study.

Third, to the best of the researchers' knowledge, this is the first study examining and supporting the moderation effect of self-choice identity revelation mode on the relationship of question-generation, peer-assessment activity and academic achievement. It was found that the predictive power of question-generation performances for academic achievement in the self-choice identity group was the greatest, and that the self-choice identity mode strengthened the relationships between question-generation and peer-assessment performances. Taking psychological safety into consideration (van Gennip, Segers, & Tillema, 2010), the results of this study, and the fact that almost all online peer-assessment systems adopt a fixed user identity mode (rather than dynamic mode that is adjustable by the user), designers of online systems should consider embedding functions that permit users to choose their own identity revelation mode.

References

Abramovich, S., & Cho, E. K. (2006). Technology as a medium for elementary preteachers' problem-posing experience in Mathematics. *Journal of Computers in Mathematics and Science Teaching, 25*(4), 309-323.

Barlow, A., & Cates, J. M. (2006). The impact of problem posing on elementary teachers' beliefs about mathematics and mathematics teaching. *School Science and Mathematics, 106*(2), 64-73.

Brindley, C., & Scoffield, S. (1998). Peer assessment in undergraduate programmers. *Teaching in Higher Education, 3*(1), 79-90.

Brown, S.I., & Walter, M. I. (2005). *The Art of Problem Posing* (3rd ed). New Jersey: Lawrence Erlbaum Associates.

Dori, Y. J., & Herscovitz, O. (1999) Question-posing capability as an alternative evaluation method: analysis of an environmental case study. *Journal of Research in Science Teaching, 36*(4), 411-30.

English, L. D. (1998). Children's problem posing within formal and in formal context. *Journal for Research in Mathematics Education, 29*(1), 83-106.

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, *70* (3), 287-322.

Gatfield, T. (1999). Examining student satisfaction with group projects and peer assessment. *Assessment & Evaluation in Higher Education, 24* (4), 365-377.

Hanrahan, S. J., & Isaacs, G. (2001). Assessing self- and peer-assessment: The students' views. *Higher Education Research & Development*, *20* (1), 53-70.

King, A. (1992). Facilitating elaborative learning through guided student-generated questioning." *Educational Psychologist 27,* 111-126.

Leung, S. S., & Wu, R. X. (1999). Problem posing with middle grades mathematics: Two real classroom examples. *Mathematics teaching in the middle school.* Reston, VA: National Council of Teachers of Mathematics.

Moses, B.M., Bjork, E., & Goldenberg, E.P. (1993). Beyond problem solving: problem posing. In S.I. Brown, and M.I. Walter. *Problem posing: reflections and applications* (pp. 178-188). NJ, Hillsdale: Lawrence Erlbaum Associates.

Purchase, H. C. (2000). Learning about interface design through peer assessment. *Assessment & Evaluation in Higher Education, 25*(4), 341-352.

Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: a review of intervention studies. *Review of Educational Research*, 66, 181–221.

Silver, E. A. (1994). On mathematical problem posing. *For the Learning of Mathematics, 14*(1), 19-28.

Topping, K. J. (2010). Methodological quandaries in studying process and outcomes in peer assessment. *Learning and Instruction, 20*(4), 339-343.

Topping, K. J. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68, 249-276.

Tsai, C. C., Liu, E. Z. F., Lin, S. S. J., & Yuan, S. M. (2001). A networked peer assessment system based on a vee heuristic. *Innovations in Education and Teaching International, 38*, 220-30.

Tsai, C. C., Lin, S. S. J. & Yuan, S. M. (2002). Developing science activities through a networked peer assessment system. *Computers & Education, 38*, 241–252.

Van Gennip, N. A. E., Segers, M., & Tillema, H. H. (2010). Peer assessment as a collaborative learning activity: The role of interpersonal variables and conceptions. *Learning and Instruction, 20* (4), 280-290.

Venables, A., & Summit, R. (2003). Enhancing scientific essay writing using peer assessment. *Innovations in Education and Teaching International, 40*, 281-290.

Vreman-de Olde, C., & de Jong, T. (2004). Student-generated assignments about electrical circuits in a computer simulation. *International Journal of Science Education* 26 (7)**,** 859-873.

Wen, M. L., & Tsai, C. C. (2006). University students' perceptions of and attitudes toward (online) peer assessment. *Higher Education, 51*, 27-44.

Wen, M. L., Tsai, C. C., & Chang, C. Y. (2006). Attitudes toward peer assessment: a comparison of the perspectives of pre-service and inservice teachers. *Innovations in Education and Teaching International, 43*, 83-92.

Wu, C. P., & Yu, F. Y. (2010). An innovation diffusion approach to online student question-generation and its effects on the relationship of perceived task value and learning approach. Paper accepted for presentation at the *18th International Conference on Computers in Education* (ICCE 2010), November 29 to December 3, Putrajaya, Malaysia.

Wu, C. P., & Yu, F. Y. (2009). Changing students' perceived value and use of learning approaches for online student-generated questions via an integrative model. *Workshop Proceedings of the International Conference on Computers in Education*, p. 30-34. November 30 to December 4. Hong Kong Institute of Education, Hong Kong.
Yu, F. Y. (2009). Scaffolding student-generated questions: Design and development of a customizable online learning system. *Computers in Human Behavior, 25* (5), 1129-1138.
Yu, F. Y., & Liu, Y. H. (2009). Creating a psychologically safe online space for a student-generated questions learning activity via different identity revelation modes. *British Journal of Educational Technology, 40* (6), 1109-1123.
Yu, F. Y., & Liu, Y. H. (2005) Potential values of incorporating multiple-choice question-construction for physics experimentation instruction. *International Journal of Science Education, 27* (11), 1319-1335.
Yu, F. Y., Liu, Y. H. & Chan, T. W. (2005). A Web-based learning system for question-posing and peer assessment. *Innovations in Education and Teaching International, 42* (4), 337-348.

Acknowledgments

This paper was funded by research grants from the National Science Council, Taiwan, ROC (NSC 96-2520-S-006-002-MY3; NSC 99-2511-S-006 -015 -MY3). The author would like to thank research assistants, Meiju Chen and Knem Chen for their assistance during the data collection process.

Using a Reflection Tool to Increase Reliability of Peer Assessments in a CSCL Environment

C.Phielix, F.J. Prins, J. Janssen, Utrecht University, P.O. Box 80.140, 3508 TC Utrecht, The Netherlands,
Email: C.Phielix@uu.nl, F.J.Prins@uu.nl, J.J.H.M.Janssen@uu.nl
P.A. Kirschner, Open University, P.O. Box 2960, 6401 DL Heerlen, The Netherlands, P.A.Kirschner@ou.nl

Abstract: To examine the reliability of students' peer assessments, two contiguous study groups used a peer assessment tool (Radar) with or without reflection tool (Reflector) in a computer-supported collaborative learning environment. Radar allows group members to assess themselves and their fellow group members on six traits related to social and cognitive behavior. Reflector stimulates group members to reflect individually and collaboratively on their past, present, and future functioning. The underlying assumption was that Radar in combination with Reflector would lead to (1) more reliable peer assessment scores, and (2) more valid perceptions of the social performance of the group. Participants were 191 second year academic students working in groups of three, four or five, on a collaborative writing task. As expected, results showed that the use of a reflection tool in a CSCL environment leads to more reliable peer assessment scores and more valid perceptions of the groups' social performance.

Introduction

Computer supported collaborative learning (CSCL) environments, though originally simple, text-based, computer mediated communication systems, have been strongly influenced by the rapid development of information and communication technology, tools and widgets (e.g., chat, video conferencing, discussion forums, group awareness widgets and shared participation tools). These applications have proven to be useful for supporting education and collaborative learning (Janssen, Erkens, Kanselaar, & Jaspers, 2007; Kreijns, Kirschner, & Jochems, 2003), leading to the design and implementation of more sophisticated CSCL environments.

Though CSCL environments have been shown to be promising educational tools and though expectations as to their value and effectiveness are high, groups learning in CSCL environments do not always reach their full potential. One of the most important reasons for this disparity between their potential and their results can be found in the social interaction between the group members, which is the key for successful collaboration (Kreijns, et al., 2003).

CSCL environments can be augmented with tools or widgets that influence social interaction (Kirschner, Strijbos, Kreijns, & Beers, 2004). Such tools, also known as 'social affordance devices', can positively affect the social performance (e.g., team development) and cognitive performance (quantity and quality of work) of a group (Kirschner, et al.). An example of such tools are self and peer assessments tools, which are increasingly applied during formative assessments and to evaluate collaborative processes during group work (Dochy, Segers, & Sluijsmans, 1999; Prins, Sluijsmans, Kirschner, & Strijbos; 2005; Strijbos & Sluijsmans, 2010). Self and peer assessment tools can provide students (and teachers) with information about group members' social and cognitive performance. However, providing group members' with information on their social and cognitive performance is not enough to positively alter their behavior (Prins, Sluijsmans, & Kirschner, 2006). Information provided by self and peer assessments are namely seldomly objective. During completion of self and peer assessments, students make many mental comparisons (Goethals, Messick, & Allison, 1991), which are selected, interpreted, and/or biased (Saavedra & Kwun, 1993). Students tend to emphasize their strengths and positive performances, and perceive weakness and negative performances as common in and caused by others (Klein, 2001; Saavedra & Kwun). Therefore, group members need to reflect individually and collaboratively (co-reflect) upon the performance of their peers and the group as a whole, before they rate themselves and their peers. Thus, in this study it is hypothesized that the use of (co-)reflection could enhance the reliability of peer assessments (e.g., Dochy, et al., 1999) and enhance behavioral change (e.g., Prins, et al., 2006). To test this hypothesis, two contiguous groups used a self and peer assessment tool with or without a reflection tool in a CSCL environment.

Self and Peer Assessment

Self assessment and peer assessment have become increasingly popular in education and CSCL (e.g., Prins, et al., 2005). Boud and Falchikov (1989) define self assessment as students making judgments about their own learning, mainly about their achievements and learning outcomes. Peer assessment can be defined as an educational arrangement where students judge a fellow student's performance qualitatively and/or quantitatively, which stimulates students to share responsibility, reflect, discuss and collaborate (Topping, 1998;

Strijbos & Sluijsmans, 2010). Somervell (1993) stresses that providing peer assessment can be seen as a part of the self assessment process, informing self assessment. Sharing self and peer assessments with others can be seen as providing information to increase group performance, therefore, self and peer assessments can be seen as a form of peer feedback (e.g., Topping, 1998; Strijbos, Narciss, & Dunnebier, 2010). Self and peer assessments can (1) provide students and teachers with a more accurate perception of students' individual behavior and performance in collaborative group work (Cheng & Warren, 2000), (2) support students in forming judgments about what can be referred to as good group behavior and high-quality performance (Topping, 1998), and (3) foster reflection on the student's own learning process and learning activities (Dochy, et al., 1999). Thus, self and peer assessments can provide students with useful information on their social and cognitive performances at both individual and group level. For this (feedback) information to be effective, students need to be challenged to reflect individually and collaboratively on their performance. Students need to ask themselves whether they understand the feedback, accept it, and determine whether it provides clues for behavioral change (Prins, et al., 2006).

Reflection

Reflection can be defined as the intellectual and affective activities individuals engage in to explore their experiences to reach new understandings and appreciations of those experiences (Boud, Keogh, & Walker, 1985). Hattie and Timperley (2007) found that for feedback to be effective, students need to answer three major questions; (1) Where am I going? (feed up), (2) How am I going? (feed back), and (3) Where to next? (feed forward). However, reflection processes are not only useful on an individual level, but also on a group level. The process in which group members collaboratively reflect on their experiences can be referred to as co-reflection. Yukawa (2006; p. 206) defined co-reflection as "a collaborative critical thinking process involving cognitive and affective interactions between two or more individuals who explore their experiences in order to reach new intersubjective understandings and appreciations". Reflection on peer feedback, thus, should make group members more aware of their own behavior, how it affects others, and whether they should alter it. This awareness allows "understanding of the activities of others, which provides a context for your own activity" (Dourish & Bellotti, 1992, p. 107). Thus, reflection can lead to new interpersonal perceptions, perspectives on experience, changes in behavior, readiness for application, and commitment to action (Boud, et al.).

Social Relations Models to Measure Reliability of Peer Assessments

In this study it is hypothesized that the use of (co-)reflection could enhance the reliability of self and peer assessments (e.g., Dochy et al., 1999) and enhance behavioral change (e.g., Prins et al., 2006). The reliability of peer assessments can be defined as the extent to which the scores for person X are consistent across all peer assessors (Bonito & Kenny, 2010). Research has shown that when students assess the performance of their peers, their assessments often dependent on one another's peer assessment (e.g., Kenny, 1994). For example, in group work, Chris's assessment of Paul is likely related to Paul's assessment of Chris. When these interdependencies are ignored, meaningful information about the interdependencies among peers is lost, and results of statistical analyses may be distorted (Bonito & Kenny; Kenny). Social Relations Models (SRM) can be used to examine these interdependencies among ratings and provide both a theoretical basis and a statistical tool (Kenny). SRM allows for variance to be partitioned into partner (assessee), actor (assessor), dyad (relationship between two assessors), and residual effects. Actor effects represent an individual's tendency to see all other group members as high or low on a particular trait, whereas partner effects reflect an individual's tendency to be seen as high or low on a particular trait by all other group members. Dyad effects represent the interaction effects of the partner and the actor at the dyadic level (i.e., does the relationship between Chris and Paul have a unique effect on the assessment even when actor and partner effects are taken into account? cf., Bonito & Kenny). In this study, SRM will be used to examine whether peer assessment scores of students with a co-reflection tool will show higher partner variances compared to students without this tool. Students with high partner variances receive more consistent (i.e., there is consensus about their cognitive or social performance) and more reliable peer assessment scores, compared to students with low partner variances.

Hypotheses

Hypothesis 1: Reflector enables to reflect upon individual and group behavior, and support students in forming judgments about what can be referred to as good group behavior and high-quality performance. Thus, students with Reflector (+Re) will perceive and receive more consistent (reliable) peer assessments scores (show higher partner variance), than students without Reflector (¬Re).

Hypothesis 2: Students with Reflector (+Re) will exhibit more realistic peer assessments scores, resulting in higher correlations between the peer assessment scores and the perceived social performance, compared to groups without Reflector (¬Re).

Hypothesis 3: Groups with Reflector (+Re) will score higher on social group performance compared to groups without Reflector (¬Re), because groups with Reflector set goals and formulate plans to enhance their social performance.

Method

Participants

Participants were 191 second-year Dutch academic Educational Science students (37 male, 154 female) with an average age of 23.64 years (*SD* = 7.16, *Min* = 19, *Max* = 55). Prior to the experiment, they were randomly assigned by the teacher to groups of three (*n* = 21), four (*n* = 160), and five (*n* = 10), and randomly assigned by the researchers to one of two conditions (see Design). Groups were heterogeneous in ability.

Table 1: Design of the study.

Condition	T1 – week 1	T2 – week 3	T3 – week 6	T4 – week 8
1. With Reflector (+Re)	Radar	Radar Reflector	Radar Reflector	Radar, Reflector, Questionnaire
2. Without Reflector (¬Re)	Radar	Radar	Radar	Radar, Questionnaire

Design

For this study, an experimental design was used with one experimental and one control condition (see Table 1). The experimental condition (*n* = 105) received a self and peer assessment tool (Radar) and a co-reflection tool (Reflector). The control condition received (*n* = 86) only Radar. During a period of 8 weeks, participants in both conditions completed the Radar four times. Additionally, from the second measurement occasion (T2: week 3), participants in the experimental condition also had to complete the Reflector.

Measures

Social behavior. Perceived group social behavior is measured by the self and peer assessments in Radar on four variables (influence, friendliness, cooperativeness, reliability). These variables are rated on a continuous scale ranging from 0 to 4 (0 = none, 4 = very high).

Cognitive behavior. Perceived group cognitive behavior is measured by the self and peer assessments in Radar on the variables 'productivity' and 'quality of contribution', rated on a continuous scale ranging from 0 to 4 (0 = none, 4 = very high).

Social performance. The perceived social performance was measured by the questionnaire at the end of the collaboration process (week 8). Four previously validated instruments (Strijbos, Martens, Jochems, & Broers, 2007) were translated into Dutch and transformed into 5-point Likert scales (1 = totally disagree, 5 = totally agree; see Table 3). The Team Development scale provides information on perceived level of group cohesion. The Group-process Satisfaction scale provides information on perceived satisfaction with general group functioning. The Intra-group Conflicts scale provides information on perceived level of conflict between group members. The Attitude towards Collaborative Problem Solving scale provides information on perceived level of group effectiveness and how group members felt about working and solving problems in a group. The 30 items in the four scales were subjected to principal component analysis. Prior to performing this analysis, the suitability of data for factor analysis was assessed. Inspection of the correlation matrix showed that all coefficients were .5 and higher. The Kaiser-Meyer-Oklin value was .73, exceeding the recommended value of .6 and Bartlett's Test of Sphericity reached statistical significance, supporting the factorability of the correlation matrix. The analysis revealed the presence of one main component with Eigen values exceeding 1, explaining 76.6% of the variance respectively. Cronbach's alpha of the composed 'Social Performance (total)' scale was .90.

Task and Procedure

Students collaborated in groups of three, four, and five, on a collaborative writing task in educational psychology. To successfully complete the collaborative writing task, each group had to write a paper about a pilot-study which they conducted over a period of eight weeks. During this period, they had to complete a self and peer assessment tool (Radar) four times, with or without a supplement co-reflection tool (Reflector). The groups used a CSCL environment called Virtual Collaborative Research Institute (VCRI; Jaspers, Broeken, & Erkens, 2004), a groupware program that supports collaborative learning on research projects and inquiry tasks. Students were instructed to make complete use of the available tools (e.g., self and peer assessment tool and reflection tool). During use of the tools, students were instructed to use the chat tool to communicate with other group members. Students received content information and definitions regarding the six traits on which they had to assess themselves and their peers. Students were told that they had eight weeks to complete the task, that it would be graded by their teacher, and that it would affect their final grade for the course. The introduction to the task stressed the importance of working together as a group and pointed out that each individual group member was responsible for the successful completion of the group task (i.e., interdependence). At the end of the final session all participants completed a 30-item questionnaire on the social performance of the group.

Instruments

Self and Peer Assessment Tool (Radar)

Radar is a self and peer assessment tool for eliciting information on group members' social and cognitive behavior visualized in a radar diagram. Radar provides students with anonymous information on how their cognitive and social behaviors are perceived by themselves, their peers, and the group as a whole with respect to specific traits found to tacitly affect how one 'rates' others (Den Brok, Brekelmans, & Wubbels, 2006). Radar provides information on six traits important for assessing behavior in groups. Four are related to social or interpersonal behavior, namely (1) influence; (2) friendliness; (3) cooperation; (4) reliability; and two are related to cognitive behavior, namely (5) productivity and (6) quality of contribution. These traits are derived from studies on interpersonal perceptions, interaction, group functioning, and group effectiveness (e.g., Den Brok, Brekelmans, & Wubbels; Kenny, 1994; Salas, Sims, & Burke, 2005). These variables, as well as the reasons for their choice, are discussed in Phielix, Prins, and Kirschner (2010) and Phielix, Prins, Kirschner, Erkens, and Jaspers (in press).

Students rate themselves and their peers on each of the six traits using a continuous scale ranging from 0 to 4 (0 = none, 4 = very high). Each range, (e.g., from a rating of 0 to 1) is divided into tenths so that every scale contained 40 points of assessment. To simplify data analysis, ratings are transformed to a 100-point scale by multiplying the ratings (0-4) by 25. Students can only access individual and average assessments of their peers after they have completed the assessment themselves. When all group members have completed their self and peer assessments, two modified radar diagrams become available in which students anonymously can (1) compare their self and received (average) peer assessments, (2) compare average peer assessment of all group members, and (3) see all personal (self and peer) assessments of their group members.

Co-reflection Tool (Reflector)

Reflector assists group members in becoming aware of their individual and group behaviour, and stimulates them to set goals and formulate plans to enhance social and cognitive group performance. Group members using Reflector individually reflect and provide information on (1) their own perspective on their personal performance (feed up), (2) differences between their self perception and the perception of their peers concerning their personal performance (feed back), (3) whether they agree with those perceptions (feed back), and (4) their individual perspective on group performance (feed up). Because group performance is determined by the individual effort of all group members, Reflector also (5) stimulates group members to collaboratively reflect (i.e., co-reflect) on group performance and reach a shared conclusion on this (feed back). Based on their shared conclusion, group members (6) set goals to improve group performance (feed forward).

The tool contained six reflective questions:

1. What is your opinion of how you functioned in the group? Give arguments to support this.
2. What differences do you see between the assessment that you received from your peers and your self assessment?
3. Why do or don't you agree with your peers concerning your assessment?
4. What is your opinion of how the group is functioning? Give arguments to support this.
5. What does the group think about its functioning in general? Discuss and formulate a conclusion shared by all the group members.
6. Set specific goals (i.e., who, what, when) to improve group performance.

The first four questions are completed individually, with completion indicated by clicking an 'Add'-button. This allows students to share their answers with the rest of the group and allows them to see the answers of the others. Students can only gain access to their peers' answers after they have added their own so as not to influence each another. The last two questions are completed in a specific frame (Co-Reflection), which allows writing a shared conclusion and formulating shared goals. Responses made by the students in Reflector are not scored or evaluated.

Results

Hypothesis 1: Reflector enables to reflect upon individual and group behavior, and support students in forming judgments about what can be referred to as good group behavior and high-quality performance. Thus, students with Reflector (+Re) will perceive and receive more consistent (reliable) peer assessments scores (show higher partner variance), than students without Reflector (¬Re).

Table 2: Proportion partner & actor variance for peer assessment scores per dependent variable per condition.

		Influence		Friendliness		Cooperation		Reliability		Productivity		Quality	
		Partner	*Actor*	*Partner*	*Actor*	*Partner*	*Actor*	*Partner*	*Actor*	*Partner*	*Actor*	*Partner*	*Actor*
+Re	T1	.20*	.35*	.09*	.68*	.11*	.64*	.06	.44*	.11	.41*	.08	.42*
	T2	.39*	.20*	.12*	.63*	.21*	.41*	.30*	.38*	.26*	.18*	.19*	.40*
	T3	.43*	.19*	.13	.49*	.29*	.34*	.43*	.18*	.40*	.22*	.38*	.21*
	T4	.28*	.35*	.04	.42*	.17*	.53*	.23*	.37*	.22*	.30*	.22	.27*
¬Re	T1	.16*	.37*	.09	.64*	.00	.60*	.05	.51*	.16*	.39*	.16*	.60*
	T2	.09	.25*	.00	.74*	.05	.70*	.06	.55*	.19*	.44*	.21*	.36*
	T3	.14*	.48*	.01	.67*	.04	.61*	.07*	.64*	.12*	.63*	.05	.65*
	T4	.20*	.46*	.05	.54*	.00	.66*	.17*	.58*	.23*	.55*	.19	.37*

* p < .05

SRM analyses were used to examine the differences in partner (assessee) and actor (assessor) variance between students with and without Reflector. Table 2 shows the partner and actor variance per dependent variable (i.e., influence, friendliness, etc.) over time per condition. As expected, for students with Reflector, all partner variances related to social behavior (influence, friendliness, cooperativeness, and reliability) are higher compared to students without Reflector. These results indicate that students with Reflector perceived (and received) more consistent peer assessments. Concerning partner variances related to cognitive behavior (e.g., productivity and quality of contribution), students with Reflector showed higher partner variances for productivity, at T2 and T3, and quality of contribution at T3. Students without Reflector showed higher partner variances for productivity at T1 and T4, as well as for quality of contribution at T1 and T2. Except for friendliness and cooperativeness at T1, and productivity at T1 and T2, all actor variances for students with Reflector are lower compared to students without Reflector. This indicates that compared to students without Reflector, peer assessments of students with Reflector are less determined by the tendency of an actor (assessor) to see all other group members as high or low on a particular trait. In contrast to students with Reflector, partner variances for students without Reflector never exceeded actor variances. For students with Reflector, at T3, partner variances were higher than actor variances for influence, reliability, productivity, and quality of contribution. Unexpectedly, for students with Reflector, partner variance for influence, cooperativeness, reliability and productivity decreases between T3 and T4. For students without Reflector, partner variance for influence, reliability, and productivity increases between T3 and T4.

Hypothesis 2: Students with Reflector (+Re) will exhibit more objective and realistic peer assessments scores, resulting in higher correlations between the peer assessment scores and the perceived social performance, compared to groups without Reflector (¬Re).

A Pearson product-moment correlation coefficient was used to test correlations between group members' average peer assessment scores and their perceived social performance at T4. Results are shown in Table 3 for students with and without Reflector.

Table 3: Correlations for average peer assessments (with and without reflector) and social performance at T4.

		Team development	Group process satisfaction	Intra group conflicts	Attitude towards CL problem solving	Social performance (total)
With Reflector (+Re)	*n*	*r*	*r*	*r*	*r*	*r*
Influence	80	.15	.31**	-.09	.13	.26*
Friendliness	80	.38**	.44**	-.28*	.19	.44**
Cooperativeness	80	.36**	.35**	-.22*	.10	.41**
Reliability	80	.24*	.19	-.16	.09	.25*
Productivity	80	.14	.17	-.12	.23*	.20
Quality of contribution	80	.14	.21	-.08	.12	.22
Without Reflector (¬Re)						
Influence	81	.15	.12	-.15	.17	.13
Friendliness	81	.04	.05	-.04	.02	.04
Cooperativeness	81	.31**	.16	-.25*	-.05	.17

Reliability	81	.25*	.21	-.23*	.05	.17
Productivity	81	.26*	.16	-.21	.04	.17
Quality of contribution	81	.28*	.20	-.21	-.02	.20

* p < .05 (2-tailed)
** p < .01 (2-tailed)

As expected, compared to students without Reflector, students with Reflector show significantly higher correlations between their ratings on influence, friendliness, cooperativeness, and their perceived social performance (in total). No significant correlations with perceived social performance (in total) were found for students without Reflector.

Hypothesis 3: Groups with Reflector (+Re) will score higher on social group performance compared to groups without Reflector (¬Re), because groups with Reflector set goals and formulate plans to enhance their social performance.

Table 4: Multilevel analyses for effects of condition on social performance scales.

	With Reflector (+Re, *n* = 89)		Without Reflector (¬Re, *n* = 86)		Comparing +Re vs. ¬Re		Chi-square
Scale	*M*	*SD*	*M*	*SD*	*β*	*SE β*	χ^2
Team development	3.60	.73	3.87	.48	-.27*	.15	3.31*
Group-process satisfaction	3.07	.41	3.14	.34	-.07	.07	.91
Intra-group conflicts	2.64	.59	2.43	.50	.21*	.12	2.85*
Attitude	3.06	.21	3.08	.18	-.02	.03	.00
Social Performance (total)	3.13	.20	3.18	.15	-.05	.04	1.60

* p < .05 (1-tailed)

Multilevel analysis was used to examine whether groups with Reflector (+Re) perceive higher social performance (i.e., better team development, higher group satisfaction, less group conflict, and more positive attitudes towards collaborative problem solving) than groups without Reflector (¬Re). Table 4 shows multilevel analyses for effects of condition on social performance scales. Unexpectedly, the significant β-value shows that groups with Reflector perceived their team as being less developed and having more intra-group conflicts, than groups without Reflector. However, no significant differences were found for total social performance, group process satisfaction, attitude towards collaborative problem solving.

Discussion & Conclusion

In this study it was hypothesized that the use of a co-reflection tool in a CSCL environment could enhance the reliability of peer assessments and enhance behavioral change. Social Relations Models (e.g., Kenny, 1994) were used to analyse the self and peer assessment data, which has never been done before in an educational or CSCL setting. Findings in this study support the assumption that supplementing a self and peer assessment tool (Radar) with a co-reflection tool (Reflector) can lead to more reliable peer assessments scores (higher partner variances). As expected, compared to students without Reflector, students with Reflector exhibited higher partner variances for their peer assessment scores on social behavior (e.g, influence, friendliness, cooperativeness and reliability), which indicate that these students perceived (and received) more consistent and reliable peer assessments. The highest partner variances for peer assessment scores on cognitive behaviour (e.g., productivity and quality of contribution) differed over time per condition. Except for productivity at T2, all actor variances for students with Reflector are lower compared to students without Reflector. This indicates that compared to students without Reflector, peer assessments of students with Reflector are less determined by the tendency of an actor (assessor) to see all other group members as high or low on a particular trait. Thus, the hypothesis that the use of a co-reflection tool in a CSCL environment could enhance the reliability of peer assessments, is accepted. Unexpectedly, for students with Reflector, partner variance decreased between T3 and T4 towards to the level of students without Reflector. A possible explanation could be that in final stage of the collaboration process students are less focused on the process (group members' behavior) and more on getting the product (paper) finished before the deadline (e.g., Aubert & Kelsey, 2003).

Findings also supported the second assumption that students with Reflector would exhibit more consistent (reliable) peer assessments scores, resulting in higher correlations between the peer assessment scores and the perceived social performance, compared to groups without Reflector. As expected, compared to students without Reflector, students with Reflector show significantly higher correlations between their peer assessment scores measured by Radar and their perceived social performance as measured by the questionnaire. No significant correlations were found for students without Reflector, indicating that the use of a co-reflection tool in a CSCL environment could enhance the validity of students' peer assessments. Thus, also the second hypothesis is accepted. An explanation for these findings could be that students without Reflector apply norm-referenced standards rather than criterion-referenced standards for assessing themselves and their peers, for instance based on prior experiences or personal beliefs. Apparently, students do not automatically reflect on a high cognitive level on their perceived and received peer assessments and need a reflection tool (i.e., Reflector) to do so (Kollar & Fischer, 2010).

Findings did not support the third assumption that the use of a co-reflection tool could enhance the social performance of the group. Unexpectedly, groups with Reflector perceived their team as being less developed and having more intra-group conflicts, than groups without Reflector. However, differences between the two conditions (with or without Reflector) are small and no significant differences were found for social performance in total. Furthermore, the perceived social performance of the students without Reflector can be argued, since these perceptions as measured by the questionnaire do not correlate with their perceived behavior (peer assessment scores) as measured by Radar.

Overall, first, SRM analyses proved to be a very useful tool to measure the variance and reliability in self and peer assessments. Second, results showed that the use of a reflection tool in a CSCL environment can lead to reliable peer assessment scores and more valid perceptions of the social performance of the group. Third, for future research on self and peer assessments in a CSCL environment, these results indicate that for self and peer assessments to be reliable, self and peer assessment tools need to be supplemented with a reflection tool.

References

Aubert, B., & Kelsey, B. (2003). Further understanding of trust and performance in virtual teams. *Small Group Research, 34*, 575–618.

Bonito, J. A. & Kenny, D. A. (2010) The measurement of reliability of social relations components from round-robin designs. *Personal Relationships, 17*, 235–251.

Boud, D., & Falchikov, N. (1989). Quantitative studies of self-assessment in higher education: a critical analysis of findings, *Higher Education, 18*, 529-549.

Boud, D., Keogh, R., & Walker, D. (1985). Promoting reflection in learning: A model. In D. Boud, R. Keogh, & D. Walker (Eds.), *Reflection: Turning experience into learning* (pp. 18–40). London: Routledge Falmer.

Brok, P. den, Brekelmans, M. & Wubbels, Th. (2006). Multilevel issues in studies using students' perceptions of learning environments: the case of the Questionnaire on Teacher Interaction. *Learning Environments Research, 9*, 199-213.

Cheng, W. & Warren, M. (2000). Making a Difference: using peers to assess individual students' contributions to a group project, *Teaching in Higher Education 5*(2), 243-255.

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education, 24*(3), 331-350.

Dourish, P., & Bellotti, V. (1992). Awareness and coordination in a shared workspace. In M. Mantel & R. Baecker (Eds.), *Proceedings of the ACM Conference on computer-supported cooperative work* (pp. 107-114). New York: ACM Press.

Goethals, G. R., Messick, D. M., & Allison, S. T. (1991). The uniqueness bias: Studies of constructive social comparison. In J. Suls & T. A. Wills (Eds.), *Social comparison research: Contemporary theory and research* (pp. 149–176). Hillsdale, NJ: Lawrence Erlbaum.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*, 81–112.

Janssen, J., Erkens, G., Kanselaar, G., & Jaspers, J. (2007). Visualization of participation: Does it contribute to successful computer-supported collaborative learning. *Computers & Education, 49*, 1037-1065.

Jaspers, J., Broeken, M., & Erkens, G. (2004). *Virtual Collaborative Research Institute (VCRI) (Version 2.0).* Utrecht: Onderwijskunde Utrecht, ICO/ISOR.

Kenny, D. A. (1994). Interpersonal perception: A social relations analysis. New York: Guilford.

Kirschner, P., Strijbos, J., Kreijns, K., & Beers, P. J. (2004). Designing electronic collaborative learning environments. *Educational Technology Research and Development*, *52*(3), 47–66.

Klein, W. M. (2001). Post hoc construction of self-performance and other performance in self-serving social comparison. *Society for Personality and Social Psychology, 27*(6), 744–754.

Kollar, I., & Fischer, F. (2010). Commentary e peer assessment as collaborative learning: a cognitive perspective. *Learning and Instruction, 20*(4), 344-348.

Kreijns, K., Kirschner, P. A. & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: a review of the research. *Computers in Human Behavior, 19,* 335-353.

Phielix, C., Prins, F. J., & Kirschner, P. A. (2010). Awareness of group performance in a CSCL-environment: Effects of peer feedback and reflection. *Computers in Human Behavior, 26*, 151-161.

Phielix, C., Prins, F. J., Kirschner, P. A., Erkens, G., & Jaspers, J. (in press). Group awareness of social and cognitive performance in a CSCL environment: Effects of a peer feedback and reflection tool.

Computers in Human Behavior, doi:10.1016/j.chb.2010.06.024

Prins, F. J., Sluijsmans, D. M. A., & Kirschner, P. A. (2006). Feedback for general practitioners in training: quality, styles, and preferences. *Advances in Health Sciences Education, 11*, 289-303.

Prins, F. J., Sluijsmans, D. M. A., Kirschner, P. A., & Strijbos, J. W. (2005). Formative peer assessment in a CSCL environment: A case study. *Assessment and Evaluation in Higher Education, 30*(4), 417–444.

Saavedra, R., & Kwun, S. K. 1993. Peer evaluation in self-managing work groups. *Journal of Applied Psychology, 78*, 450-462.

Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a "Big Five" in teamwork? *Small Group Research, 36*, 555-599.

Somervell, H. (1993). Issues in assessment, enterprise and higher education: the case for self-, peer and collaborative assessment, *Assessment and Evaluation in Higher Education, 18*, 221-233.

Strijbos, J.W., Martens, R. L., Jochems, W. M. G., & Broers, N. J. (2007). The effect of functional roles on perceived group efficiency during computer-supported collaborative learning: a matter of triangulation. *Computers in Human Behavior*, *23*, 353–380.

Strijbos, J. W., Narciss, S., & Dünnebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: are they critical for feedback perceptions and efficiency? *Learning and Instruction, 20*(4), 291-303.

Strijbos, J. W., & Sluijsmans, D. M. A. (2010). Guest editorial - Unravelling peer assessment: methodological, functional, and conceptual developments. *Learning and Instruction, 20*(4), 265-269.

Topping, K. (1998). Peer assessment between students of colleges and universtities. *Review of Educational Research, 68(3),* 249-276.

Yukawa, J. (2006). Co-reflection in online learning: Collaborative critical thinking as narrative. *International Journal of Computer-Supported Collaborative Learning, 1*(2), 203-228.

Adaptive Support for CSCL: Is it Feedback Relevance or Increased Student Accountability that Matters?

Erin Walker, Carnegie Mellon University, Pittsburgh, USA, erinawalker@gmail.com
Nikol Rummel, Ruhr-Universität Bochum, Bochum, Germany, nikol.rummel@rub.de
Kenneth R. Koedinger, Carnegie Mellon University, Pittsburgh, USA, koedinger@cmu.edu

Abstract: While fixed CSCL support approaches such as collaboration scripts have been shown to improve domain learning, adaptive support that varies based on student actions may be more effective. In this paper, we discuss the Adaptive Peer Tutoring Assistant (*APTA*), an adaptive support system for computer-mediated peer tutoring in high-school Algebra. We conducted an after-school study with 122 participants where we compared *APTA* to two fixed support conditions: one where we told students support was adaptive when it was not, and one where we told students support was fixed. These manipulations explored two hypotheses: Adaptive support is effective because it is relevant to student behaviors, and support that is perceived to be adaptive is effective because it makes students feel accountable for their actions. *APTA* showed better effects on tutor and tutee learning compared to the other conditions, suggesting that the more relevant the support, the more beneficial it will be.

Introduction

Collaborative activities in the classroom can yield positive learning outcomes through the social construction of knowledge (Schoenfeld, 1992), but students need to engage in beneficial reflective and elaborative processes, and generally do not do so without assistance (Lou, Abrami, & d'Apollonia, 2001). Collaboration can be supported using *scripts*, where interaction is structured by assigning students roles and activities to follow in their interaction (e.g., Fischer, Kollar, Mandl, & Haake, 2007). For example, in the reciprocal teaching script for reading comprehension by Palincsar and Brown (1984), students alternate between the roles of tutor and tutee, and engage in a sequence of elaborative activities involving summarizing, questioning, clarifying and predicting. Other script approaches structure student dialogue on a fine-grained level by providing questions (e.g. "What would happen if ...") or sentence starters (e.g. "It was found that ...") that students apply during their collaboration (Kollar, Fischer, & Slotta, 2005). While scripts have been successful, they may not be maximally effective at improving student collaboration: providing too much support for students who do not need it or too little support for students who do (e.g., Kollar et al., 2005). Further, they might have drawbacks because they overstructure student interaction, limiting student control, and potentially reducing feelings of being good collaborators and desire to be good collaborators (Dillenbourg, 2002).

A promising new method for facilitating computer-supported collaborative activities in the classroom is by providing students with *adaptive* collaborative learning support (ACLS), where interaction is analyzed as it occurs and then support is provided tailored to the needs of individual collaborators. The few adaptive support systems that have been evaluated with students have shown benefits compared to fixed support (Baghaei, Mitrovic, & Irwin, 2007; Kumar, Rosé, Wang, & Joshi, & Robinson, 2007). However, because of the technical challenges in developing robust ACLS systems, these evaluations are difficult to conduct, and it is still unclear under what conditions adaptive support is more effective than fixed support. In the study presented in this paper, we explore two mechanisms for the potential effectiveness of adaptive support, illustrated in Figure 1. First, we hypothesize that adaptive support in collaborative learning is effective for the same reasons as in individual learning: it is more relevant than fixed support (*H1: Relevance Hypothesis*). Students might benefit from being able to immediately apply the support to their interaction, leading them to improve their collaboration and their domain learning (Rummel & Weinberger, 2008). Additionally, because support only appears when it is relevant it avoids the overstructuring problem of fixed support, potentially increasing student feelings of being good collaborators (*perceived collaboration efficacy*). A second hypothesis is that students who *believe* that support is adaptive may feel more accountable for their collaborative actions, and thus be more motivated to collaborate effectively (*H2: Accountability Hypothesis*). Accountability for one's partner's outcomes tends to be an important motivational force in collaboration (e.g., Slavin, 1996), and it is plausible to assume that if students believe the computer is responding to their actions, they may feel an increased sense of responsibility for those actions. If this hypothesis is correct, we would see benefits of collaboration on both student goals to be effective collaborators (*collaboration goal orientation*) and domain learning. Distinguishing between these two hypotheses is important because if *H2* is true, it suggests that it may not be necessary to develop sophisticated adaptive systems, but simply to develop systems that can convincingly pretend to be adaptive.

To examine these two hypotheses, we developed a system to adaptively support a reciprocal peer tutoring activity for high school algebra, building on a successful individual intelligent tutoring system, the Cognitive Tutor Algebra *(CTA*; Koedinger, Anderson, Hadley, & Mark, 1997*)*. We conducted a study where we

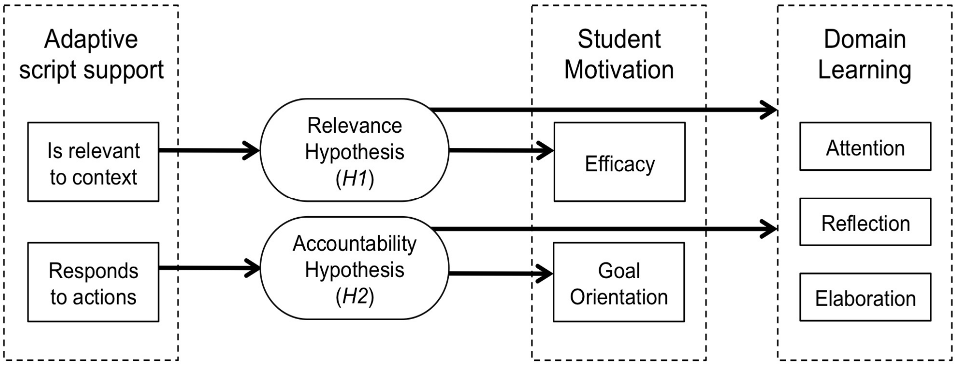


Figure 1. Adaptive support may influence domain learning and motivation by being more relevant to student needs (*H1*), which should enhance self-efficacy and learning, or by encouraging greater accountability (*H2*), which should enhance goal orientation and learning.

compared an adaptive to a fixed support system, by varying the adaptivity of reflective prompts given to peer tutors to improve their help-giving. In order to differentiate between the two hypotheses, and determine whether adaptive support could be solely effective through the *accountability hypothesis (H2),* we included a third condition where we told students support was adaptive, when it in fact was fixed. Student belief that support is adaptive may lead to benefits that can be explained by *H2* but not *H1*. In this paper, we first discuss the reciprocal peer tutoring context for the adaptive support. We then describe the adaptive support system we have developed for reciprocal peer tutoring. Finally, we discuss the study methodology and its results.

Context: Help-Giving in Reciprocal Peer Tutoring

Help-giving is an important part of many collaborative activities, and is a key element of the productive interactions identified by Johnson and Johnson (1990) that contribute to learning from collaboration. The act of giving help has been demonstrated to improve learning of both the help giver and the receiver (see Ploetzner, Dillenbourg, Preier, & Traum, 1999). In giving help, even novice students benefit through *reflective* and *elaborative* processes; they reflect on their peer's error and then construct a relevant explanation, elaborating on their existing knowledge and generating new knowledge (Roscoe & Chi, 2007). In turn, students benefit from receiving good help; that is, help that arrives when they reach an impasse, allows them to self-explain, and, if necessary, provides an explanation that is conceptual, targets their misconceptions, and is correct (Webb & Mastergeorge, 2003). Unfortunately, most students do not exhibit good help-giving behaviors spontaneously (Roscoe & Chi, 2007). Specifically, students are often more inclined to give each other instrumental help (e.g., "subtract *x*"). They rarely provide conceptual, elaborated help that explains why, in addition to what, and references domain concepts (e.g., "subtract *x* to move it to the other side"). This tendency decreases the likelihood that either student benefits from the interaction (Webb & Mastergeorge, 2003). Therefore, promoting conceptual help-giving is a major focus of peer tutoring support. It is one criteria of good help for the receiver, and an indicator that the help-giver engaged in elaborative processes.

One technique for facilitating student help-giving is by employing a reciprocal schema, where students are given different information and take turns helping each other using the information they received (Dillenbourg & Jermann, 2007). As part of their role, helpers must monitor their partner's problem solving and offer appropriate explanations when needed. Examples of this class of collaborative activities are reciprocal teaching by Palincsar and Brown (1984), mutual peer tutoring by King, Staffieri, and Adelgais (1998), and reciprocal peer tutoring by Fantuzzo, Riggio, Connelly, and Dimeff (1989). Reciprocal learning settings have been successful at increasing student learning in classroom environments compared to individual and unstructured controls (Fantuzzo et al., 1989; King et al., 1998; Fuchs et al., 1997), but only when student activities are appropriately supported. For example, Fuchs et al. (1997) trained students to deliver conceptual mathematical explanations and give elaborated help, and showed that their mathematical learning was significantly better than elaborated help training alone or an individual learning control. There is reason to believe that adaptive support would be effective in this context, for the same reasons that it may be effective in other collaborative contexts. Adaptive support might provide relevant assistance at moments when it most needed, increasing student reflective and elaborative processes, and thus improving domain learning. Additionally, motivation plays an important part of peer tutoring. Peer tutors who feel accountable for their partner's learning tend to attend more to the subject material and learn more (e.g., Biswas et al., 2005), and peer tutors who feel like good tutors tend to try harder at the activity (Medway & Baron, 1977). Adaptive support in this context might have an amplified effect on student motivation, enabling us to investigate both the *relevance* and *accountability hypotheses*.

System: The Adaptive Peer Tutoring Assistant

Basic Peer Tutoring Script

To investigate the effects of adaptive collaboration support, we constructed a computer-supported collaborative learning system called the Adaptive Peer Tutoring Assistant (*APTA*). In *APTA*, dyads of students work on literal equation problems where they are given an equation like "$ax + by = c$" and a prompt like, "Solve for x". Students are seated at different computers, and at any given time, one student is the peer tutor and the other is the tutee. Tutees can perform operations on the equation with a menu-based interaction used in the common, individual version of the Cognitive Tutor Algebra (*CTA*). Using the menus, students can select operations like "subtract from both sides", and then type in the term they would like to subtract. For some problems, the computer then performs the result of the operation and displays it on the screen; for others, the tutee must type in the result of the operation themselves. The peer tutors can see the tutee's actions on their computer screen, but are not able to perform actions in the problem themselves (#5 in Figure 2). Instead, the peer tutor can mark the tutee's actions (#6 in Figure 2), and adjust tutee skill assessments in the skillometer window (#1, Figure 2). Students can discuss the problem in a chat window (#4 in Figure 2). We intended that in constructing help for tutees in the chat window, peer tutors would engage in elaborative aspects of tutoring, generating knowledge and integrating it with existing knowledge. To facilitate the discussion in the chat window, we included a common form of fixed scaffolding: sentence classifiers. This form of fixed scaffolding is thought to be pedagogically beneficial by making positive collaborative actions explicit in the interface and encouraging students to consider the type of utterance they wish to make (Weinberger, Ertl, Fischer, & Mandl, 2005). We asked peer tutors to label their utterances using one of four classifiers: "ask why", "explain why wrong", "give hint", and "explain what next" (#8 in Figure 2). Students had to select a classifier before they typed in an utterance, but they could also choose to click a neutral classifier ("other"). For example, if students wanted to give a hint, they could click "give hint" and then type "subtract x". Their utterance would then appear as "tutor hints: subtract x" to both students in the chat window. By making those behaviors explicit in the interface, we hoped to encourage students to put more consideration into what they said and why, facilitating them in constructing more conceptual help.

In the basic version of *APTA*, we provided tutors with adaptive domain assistance that supported them in reflecting on tutee errors. We intended that this assistance have the additional benefit of ensuring that the

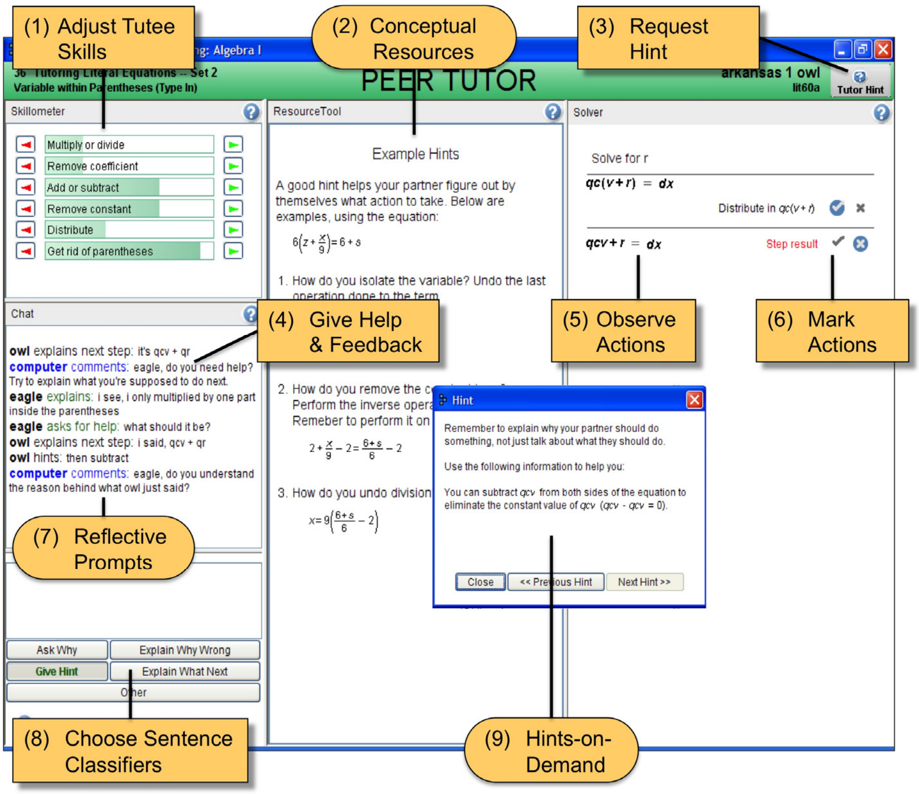


Figure 2. Peer tutor's interface in *APTA*. Tutors observe and mark tutee actions, assess tutee skills, and provided tutees with explanations in the chat window. They receive domain hints and feedback, and adaptive reflective prompts to help them improve their collaboration.

tutee received more correct help than they otherwise would have. This assistance was provided in two cases. First, the peer tutor could request a hint from the CTA and relay it to the tutee. Second, if the peer tutor marked something incorrectly in the interface (e.g., they marked a wrong step by the tutee correct), the intelligent tutor would highlight the answer in the interface, and present the peer tutor with an error message. Hints and error messages were composed of a collaborative component ("Remember to explain why your partner should do something, not just what they should do"), and the cognitive component that the tutee would have originally received had they been using the CTA individually ("You can subtract *qcv* from both sides of the equation to eliminate the constant value of *qcv* [*qcv* – *qcv* = 0]"; see #9, Figure 2). If the tutor clicks next hint, both components become more specific, until the cognitive component reveals the answer to the tutor.

Reflective Prompts: Study Manipulation

Our experimental manipulation varied the adaptivity of reflective prompts students receive while collaborating. These reflective prompts appear simultaneously to both students in the chat window and target peer tutor help-giving skills that need improvement (#7 in Figure 2). For example, novice tutors may give instrumental hints like "then subtract" rather than conceptual hints like "to get rid of *qcv,* you need to perform the inverse operation on that side of the equation." When tutors are detected to be giving instrumental hints, the computer uses an assessment of the tutor's help-giving skill to say in the chat window (visible to both students), "Tutor, think about the last help you gave. Why did you say that? Can you explain more?" This utterance is designed to get both students reflecting on the domain concepts behind the next step, and to remind the tutor that help should explain why in addition to what. Prompts are addressed to the peer tutor (e.g., "Tutor, can you explain your partner's mistake?"), and are adaptively selected based on the computer assessment of help-giving skills. For example, as use of the sentence classifiers is an integral component of our ability to assess peer tutor utterances, as well as having potential benefit for the students, when students fail to use the sentence classifiers, they receive prompts suggesting that they should do so (e.g., "The buttons underneath the chat [e.g., "Give Hint"] can help you let your partner know what you're doing"). Students also receive encouragement when they display a particular help-giving skill (e.g., "Good work! Explaining what your partner did wrong can help them not make the same mistake on future problems"). The prompts contain both praise and hedges, such that the computer's voice does not publicly threaten the peer tutor's voice. Only one reflective prompt was given at a time, and parameters were tuned so that students received an average of one prompt for every three peer tutor utterances. There were several different prompts for any given situation, so students rarely received the same prompt twice.

We built a model for good peer tutoring which assessed whether students displayed four help-giving skills: help in response to tutee errors and requests, help that targets tutee misconceptions, help that is conceptual and elaborated, and the use of sentence classifiers to give help. This assessment is the basis for providing students with reflective prompts. Our main focus was supporting peer tutors in giving conceptual elaborated help to benefit their own learning. For example, by encouraging peer tutors to target tutee misconceptions, we hoped to lead them to reflect and elaborate more on the concepts involved in solving the problem. To assess peer tutor performance, the model uses a combination of several inputs. First, it uses *CTA* domain models to access the problem-solving context (e.g., it could tell if tutees had just made an error). Next, it uses student interface actions, including tutor self-classifications of chat actions as prompts, error feedback, hints, or explanations, to determine what the students' intentions were when giving help. Finally, it used Taghelper (Rosé et al., 2008) to automatically determine whether students were giving help, whether the help targeted the next problem step or the previous problem step, and whether the help was conceptual. Based on a combination of these three channels, we used a simple model composed of 15 rules to assess each peer tutor action taken, and used Bayesian knowledge tracing (Corbett & Anderson, 1995) to update a running assessment of peer tutor mastery of the four help-giving skills. For example, if the peer tutor clicked the "give hint" classifier and typed "subtract *x*", the system would classify the utterance as nonconceptual help on the next problem step. The system would then access the *CTA* problem-solving context, and might see that the tutee had recently made an error. The system would consider the peer tutor utterance to be suboptimal collaboration because the peer tutor did not give feedback on the error, and the "target misconceptions" skill assessment would be decreased. If, after any peer tutor action, a skill fell within a predefined threshold for that skill, students were given a reflective prompt targeting that skill. In this particular example, the peer tutor may be told: "Tutor, is there anything your partner doesn't understand right now about the problem?" These prompts are adaptive both with respect to their timing and with respect to their content.

Method

The goals of this study were to investigate the potential beneficial effects of adaptive support on collaboration and learning in a computer-supported peer tutoring setting and to explore two potential mechanisms for these effects. Thus, we compared three conditions:

1) Students received adaptive support and were told it was adaptive (*real adaptive condition*)

2) Students received fixed support and were told it was adaptive (*told adaptive condition*)
3) Students received fixed support and were told it was fixed (*real fixed condition*)

We deployed the three versions of our system in an after-school lab study to examine the influence of the actual and perceived adaptive support on student learning and motivation. Following the *accountability hypothesis* (*H2*), we believed that in the conditions where students perceived support as adaptive (the *real* and *told adaptive conditions*), they would be more motivated to collaborate effectively, and would learn more than the students who perceived the support as random. Following the *relevance hypothesis* (*H1*), we believed that in the condition where support was actually adaptive (only the *real adaptive condition*), students would receive feedback on their help only at moments when they could apply it, and thus would learn more than other students and feel more positively about their tutoring.

Participants

Participants were 130 high-school students (49 males, 81 females) from one high school, ranging from 7th to 12th grade, and currently enrolled in Algebra 1 (46 students), Geometry (49 students), or Algebra 2 (35 students). While the literal equation solving unit was one that all students had completed, the teacher we were working with identified it as a challenging unit for the students, and, in fact, many students did not remember seeing the material before. The study was run at the high school, either immediately after school or on Saturdays. All students were paid 30 dollars for their participation. Students participated in sessions of up to 9 students at a time (*M* group size = 7.41, *SD* = 1.35). Each session was randomly assigned to one of the three conditions, and then within each pair students were randomly assigned to the role of tutee or tutor. Students came with partners that they had chosen, except in the case of 4 students assigned to their partners by the researchers. For ease of scheduling, we sometimes assigned an extra student to a given session (in case somebody did not show up at the assigned time). There were 8 students who worked alone over the course of the session. Thus, a total of 122 students were included in the analysis.

Procedure

This study took place over 3 hours. Students received a brief 5-minute introduction to the study, and then took a 20-minute domain pretest. Next, students spent 20 minutes in a preparation phase, working individually using the *CTA*. Students solved problems involving literal equations where the variable terms were on the same side of the equation (e.g., $ax + bx = cy + dz$; solve for x). Students then spent 30 minutes in the tutoring phase, with one student tutoring another student on problems where the variable terms were on both sides of the equation (e.g., $ax + cy = bx + dz$; solve for x). Students took up to 10 minutes to answer several survey questions on their motivational state, and then spent another 30 minutes in the tutoring phase. At this point, students took a 15-minute break, and then took a 20-minute domain posttest, again consisting of a 10-minute conceptual component and 10-minute procedural component. Students concluded the study by tutoring without support for 25 minutes, and answering demographic questions.

In the tutoring phase, we implemented two between-subjects manipulations where we varied whether students received adaptive support and whether they thought the support was adaptive. As our first manipulation, we varied the adaptivity of the reflective prompts students received. To implement the comparison conditions where students received fixed support, we gave students pseudo-random prompts that ensured that the timing and content of the prompts was not contingent on student actions. Every time students would have received a reflective prompt were they in the adaptive condition, they did not receive a prompt in the fixed condition. However, we ensured that they received a prompt within the next three turns, thus yoking the fixed prompt to the adaptive prompt. We randomly selected the content of the prompt, with one exception: we did not choose content that would have been parallel to the yoked adaptive prompt. Thus, we tried to avoid cases where the fixed support prompts were accidentally relevant (according to our model of adaptivity), making the comparison between the adaptive and fixed support more controlled. For the second manipulation, prior to the tutoring phase, we gave students instructions that told them that the support was either adaptive or fixed. The adaptive instructions were as follows: "The computer will watch you tutor, and give you targeted advice when you need it based on how well you tutor. Both you and your partner will see the help in the chat." The fixed instructions were as follows: "From time to time, the computer will give you a general tip chosen randomly from advice on good collaboration. Both you and your partner will see the help in the chat." As students began to use the tutoring system, they were given further instruction on how to use the system, including instructions on how to indicate how they felt about the reflective prompts using "like" and "dislike" widgets. To reaffirm the experimental manipulation, students in the *real* and *told adaptive* conditions were told: "We will use that information to improve the computer's ability to track what you're doing and give you advice you can use." Students in the *real fixed* condition were told: "We will use that information to describe which pieces of advice can go into the pool of advice we randomly select from."

Measures

To assess students' individual *domain learning* we used counterbalanced pretests and posttests, each containing 7 conceptual items and 5 procedural items. Tests were approved by the coordinating classroom teacher, and were administered on paper. We scored answers on these tests by marking whether students were correct or incorrect on each item, and then summing the scores. We further assessed student motivational state in two ways. First, we included several items assessing *perceived collaboration efficacy* relating to how positively students perceived themselves in the interaction (e.g., for the tutor: "I think I was a good tutor"), and how positively they perceived their partner (e.g., for the tutee: "I think my partner learned a lot from being a tutor"). Second, we adapted individual learning orientation questionnaires (Elliot & McGregor, 2001) to assess peer tutor mastery and performance goals for being a good tutor (e.g., "While tutoring, I was worried that I might not learn enough about tutoring", "While tutoring, my goal was to show my partner I was a good tutor"), and tutee mastery and performance goals for helping their partner be a good tutor (e.g., "While being tutored, I wanted my partner to understand how to tutor", "While being tutored, it was important for me that my partner look like a good tutor"). We called this measure *collaboration goal orientation.* For both motivation measures, scores were averaged across all relevant items. Finally, as a manipulation check, we assessed *perceived adaptivity,* with five items asking students how adaptive they thought the system was (e.g., for the tutor: "The computer gave advice at times when it was useful.") and how positively they perceived the system (e.g., for the tutee: "The advice the computer gave improved how well my partner tutored me.").

Results

Domain Learning

We conducted a two-way (condition x role) ANCOVA, controlling for pretest, with posttest as the dependent variable. Pretest score was significantly predictive of posttest score ($F[1,115]=120.43$, $p < 0.001$; see Table 1). There was a significant effect of condition on posttest ($F[2,115] = 4.20$, $p = 0.017$, $eta^2 = 0.068$), indicating that the adaptiveness of support had a positive effect on student posttest performance. A planned comparison of the effects of receiving real adaptive support revealed that it indeed had a significant effect ($F[1,115] = 7.47$, $p = 0.007$), while a planned comparison of the effects of receiving support that students were told was adaptive revealed that this manipulation did not have a significant effect ($F[1,115]=0.393$, $p = 0.532$). These results support the *relevance* (*H1*) but not the *accountability* (*H2*) hypothesis, suggesting that real adaptive support had a more beneficial effect than fixed support, even if students were told that the support was adaptive.

Interestingly, while the effect of role on posttest was not significant ($F[1,115] = 0.751$, $p = 0.338$), there was a significant interaction effect between condition and role ($F[1,115] = 3.334$, $p = 0.039$, $eta^2 = 0.055$). Applying the planned comparisons based on *H1* and *H2* to the interaction effect revealed that while the effects of real adaptivity did not differ across the two roles ($F[1,115] = 2.660$, $p = 0.106$), told adaptivity had differential effects on peer tutors and tutees ($F[1,115] = 6.561$, $p = 0.012$). Inspecting student learning across role and condition (see Table 1) revealed that peer tutors benefit more from the told adaptive condition than the real fixed condition, but tutees benefit more from the real fixed condition than the told adaptive condition. The perception of adaptivity may have an effect on peer tutor feelings of accountability and thus may positively influence their learning. However, the perception that the tutoring advice is relevant when it is not may impede the tutoring abilities of the peer tutor and thus may lead to less tutee learning than in the real fixed condition.

Table 1. Pretest and posttest scores for the tutee and tutor. Scores represent percent correct.

	Real Fixed		Told Adaptive		Real Adaptive	
	Tutor	Tutee	Tutor	Tutee	Tutor	Tutee
Pretest Score	0.29 (0.15)	0.23 (0.16)	0.24 (0.12)	0.28 (0.18)	0.27 (0.16)	0.28 (0.15)
Posttest Score	0.28 (0.18)	0.33 (0.21)	0.27 (0.16)	0.29 (0.14)	0.39 (0.17)	0.36 (0.21)

Tutoring Efficacy and Goal Orientation

We then investigated the motivational effects of the manipulation. We assessed how positively students felt about their and their partner's tutoring abilities in the *perceived collaboration efficacy* measure, using a two-way (condition x role) ANOVA (see Table 2, row 1). We found that condition significantly affected student positive feelings of perceived collaboration efficacy ($F[2,102] = 5.58$, $p = 0.005$), as did role ($F[1,102] = 5.10$, $p = 0.026$). There was no significant interaction of condition and role ($F[2,102] = 0.542$, $p = 0.583$). We then looked at the effects of condition on tutoring mastery and performance orientation using the *collaboration goal orientation* measure (see Table 2, rows 2 and 3). Condition did not have a significant effect on either variable (mastery orientation: $F[2,99] = 0.501$, $p = 0.607$; performance orientation: $F[2,99] = 0.679$; $p = 0.510$).

Table 2. Motivational effects and manipulation check. Standard deviations are presented in parentheses. Scores are on a 7-point Likert scale, with 7 being the most positive response.

	Real Fixed		Told Adaptive		Real Adaptive	
	Tutor	Tutee	Tutor	Tutee	Tutor	Tutee
Collaboration efficacy	4.53 (1.28)	5.35 (1.37)	4.68 (1.86)	4.90 (1.39)	5.43 (1.03)	6.14 (0.67)
Mastery orientation	4.93 (1.10)	5.15 (1.24)	4.94 (1.05)	5.10 (1.17)	5.28 (1.10)	5.46 (1.52)
Performance orientation	4.77 (1.17)	5.08 (1.01)	4.78 (1.25)	4.72 (1.59)	4.85 (1.05)	4.82 (1.25)
Perceived adaptivity	4.75 (1.60)	4.79 (1.56)	5.06 (1.05)	4.54 (1.30)	4.84 (1.30)	5.23 (1.25)

Perceived Adaptivity

As a manipulation check, we evaluated how adaptive and effective students thought the system was (*perceived adaptivity*; see Table 2, row 4). In a two-way (condition x role) ANOVA, there were no significant effects of condition on perceived adaptivity ($F[2, 96] = 1.046$, p = 0.355), no significant effects of role ($F[1,96] = 0.00$, $p = 0.992$), and no interaction ($F[2,96] = 1.741$, $p = 0.181$).

Discussion and Conclusions

In this paper, we described *APTA*, a system for adaptively supporting peer tutoring, and conducted a study where we compared a condition employing *APTA* to two conditions employing a fixed support system. The results suggested that compared to fixed support, adaptive support improved the domain learning of peer tutors and tutees. This result adds to the small list of studies demonstrating the effectiveness of adaptive collaborative support at improving learning (e.g., Kumar et al., 2007). We also found that students who received adaptive support thought they were better collaborators than their fixed support peers. As there is a positive link between efficacy and motivation, this finding suggests that adaptive reflective prompts may help overcome motivational drawbacks of fixed support systems (e.g., Dillenbourg, 2002).

One of the central questions of this research agenda is: how adaptive does support need to be? The effectiveness of the real adaptive support compared to the told adaptive support provided evidence for the *relevance hypothesis* (*H1*; the top path in Figure 1). As the fixed support conditions received approximately the same collaborative support content as the adaptive support condition, we can conclude that it was the adaptive presentation of the content at appropriate moments that enhanced student learning and motivation. We also tested whether by simply telling students that support was adaptive, even when it was not, students might learn more from the collaboration (*accountability hypothesis*; *H2*; the bottom path in Figure 1). This hypothesis was not confirmed. Peer tutors may have received some benefit from the perception of adaptivity, but tutees performed worst in this condition. Moreover, the adaptivity of support had a significant effect on collaboration efficacy but not goal orientation, further supporting the *relevance hypothesis* over the *accountability hypothesis.* Despite these results, we found no differences between conditions in student reports of how adaptive they found the system. It is interesting that although students in the real adaptive condition did not appear, on the survey measure, to perceive the system as more adaptive, they still received beneficial effects of adaptivity.

This study tried to identify guidelines for supporting collaborative learning using controlled experimentation in a school context, and further focused on uncovering the underlying mechanism by manipulating particular aspects of adaptive support. Theoretical progress in understanding the potential of adaptive support for collaborative learning will be enhanced by further investigations of alternative mechanisms for how, when, and why particular forms of adaptive support are effective. In general, the positive learning and motivational results relating to real adaptive support encourage the continuation of research into adaptive support for collaboration, suggesting that the time and effort necessary to develop this support is worthwhile.

References

Baghaei, N., Mitrovic, A., & Irwin, W. (2007). Supporting Collaborative Learning and Problem Solving in a Constraint-based CSCL Environment for UML Class Diagrams. *International Journal of Computer-Supported Collaborative Learning*, *2* (2-3), 159-190.

Biswas, G., Leelawong, K., Schwartz, D., Vye, N. & The Teachable Agents Group at Vanderbilt (2005). Learning By Teaching: A New Agent Paradigm for Educational Software. *Applied Artificial Intelligence, 19*, 363-392.

Corbett, A.T., Anderson, J.R. (1995) Knowledge Tracing: Modeling the Acquisition of Procedural Knowledge. *User Modeling and User-Adapted Interaction, 4*, 253-278.

Dillenbourg, P. (2002). Over-scripting CSCL: The risk of blending collaborative learning with instructional design. In Kirschner, P. A. (Ed.), Three worlds of CSCL: Can we support CSCL? 61-91.

Dillenbourg, P., & Jermann, J. (2007). Designing integrative scripts. In Scripting Computer-Supported Collaborative Learning - Cognitive, Computational, and Educational Perspectives, Computer-Supported Collaborative Learning Series, pages 275-301. Springer, New York, 2007.

Elliot, A. J., & McGregor, H. A. (2001). A 2 x 2 achievement goal framework. Journal of *Personality and Social Psychology, 80*, 501-519.

Fantuzzo, J. W., Riggio, R. E., Connelly, S., & Dimeff, L. A. (1989). Effects of reciprocal peer tutoring on academic achievement and psychological adjustment: A component analysis. *Journal of Educational Psychology, 81*(2), 173-177.

Fischer, F., Kollar, I., Mandl, H., & Haake, J. (2007). Scripting Computer-Supported Collaborative Learning – Cognitive, Computational, and Educational Perspectives. *Computer-Supported Collaborative Learning Series*, New York: Springer.

Fuchs, L., Fuchs, D., Hamlett, C., Phillips, N., Karns, K., & Dutka, S. (1997). Enhancing students' helping behavior during peer-mediated instruction with conceptual mathematical explanations. *The Elementary School Journal, 97*(3), 223-249.

Johnson, D. W. & Johnson, R. T. (1990). Cooperative learning and achievement. In S. Sharan (Ed.), *Cooperative learning: Theory and research* (pp. 23-37). NY: Praeger.

King, A., Staffieri, A., & Adelgais, A. (1998). Mutual peer tutoring: Effects of structuring tutorial interaction to scaffold peer learning. *Journal of Educational Psychology, 90*, 134-152.

Koedinger, K., Anderson, J., Hadley, W., & Mark, M. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, 8, 30-43.

Kollar, I., Fischer, F., & Slotta, J. D. (2005). Internal and external collaboration scripts in web-based science learning at schools. In T. Koschmann, D. Suthers, & T.-W. Chan (Eds.), The next 10 years*! Proceedings of the International Conference on Computer Support for Collaborative Learning* 2005 (pp. 331-340). Mahwah, NJ: Lawrence Erlbaum Associates.

Kumar, R., Rosé, C. P., Wang, Y. C., Joshi, M., Robinson, A. (2007). Tutorial dialogue as adaptive collaborative learning support. In R. Luckin, K. R. Koedinger, & J. Greer (Eds.) *Proceedings of Artificial Intelligence in Education* (pp. 383-390). IOS Press.

Lou, Y., Abrami, P. C., d'Apollonia S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of Educational Research, 71*(3), 449-521.

Medway, F. & Baron, R. Locus of control and tutors' instructional style. *Contemporary Educational Psychology, 2*, 298-310 (1997).

Palincsar, A.S., & Brown, A.L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1*(2), 117-175.

Ploetzner, R., Dillenbourg, P., Preier, M., & Traum, D. (1999). Learning by explaining to oneself and to others. In P. Dillenbourg (Ed.), *Collaborative Learning: Cognitive and Computational Approaches* (pp. 103 – 121). Elsevier Science Publishers.

Roscoe, R. D. & Chi, M. (2007) Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research. 77*(4), 534-574.

Rosé, C. P., Wang, Y.C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., Fischer, F. (2008). Analyzing Collaborative Learning Processes Automatically: Exploiting the Advances of Computational Linguistics in Computer-Supported Collaborative Learning. *International Journal of Computer-Supported Collaborative Learning, 3*(3), 237-271.

Rummel, N. & Weinberger, A. (2008). New challenges in CSCL: Towards adaptive script support. In G. Kanselaar, V. Jonker, P.A. Kirschner, & F. Prins, (Eds.), International perspectives of the learning sciences: Cre8ing a learning world. *Proceedings of the Eighth International Conference of the Learning Sciences (ICLS 2008), Vol 3* (pp. 338-345). International Society of the Learning Sciences.

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem-solving, metacognition, and sense making in mathematics. In D. Grouws (Ed.), *Handbook for research on mathematics teaching and learning* (pp. 334–370). New York: Macmillan.

Slavin, R. E. (1996). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary Educational Psychology, 21*, 43-69.

Webb, N. M., & Mastergeorge, A. (2003). Promoting effective helping behavior in peer-directed groups. *International Journal of Educational Research*, *39*, 73-97.

Weinberger, A., Ertl, B., Fischer, F., & Mandl, H. (2005). Epistemic and social scripts in computer-supported collaborative learning. Instructional Science, 33(1), 1-30.

Acknowledgments

This research is supported by the Pittsburgh Science of Learning Center, NSF Grant #SBE-0836012. Thanks to Carolyn Rosé, and the mathematics coordinator who made this study possible.

Mind the Gap: Differences between the Aspirational and the Actual in an Online Community of Learners

Karen Brennan, MIT Media Lab, 77 Massachusetts Ave, E14-445C, Cambridge, MA, 02139, kbrennan@media.mit.edu

Abstract: This paper explores the differences between the *aspirational* (what designers desire) and the *actual* (what participants do) in an online learning community of young people, using the Scratch website as a case study. Interviews with members of the MIT Scratch Team and two years of ethnographic observation of the site inform the discussion of these differences. The paper presents four tensions that arise between the aspirations for this community of learners and the realities of actual participation: (1) enabling exploration of the personal vs. pitfalls of the personal, (2) fostering meaningful interactions with others vs. lack of respect, (3) supporting diverse creative expression vs. conflicts in interests, and (4) encouraging a sense of group belonging vs. seeking attention.

Scratch: An Online "Community" of Learners

Scratch (http://scratch.mit.edu) is a programming environment that makes it easy to create interactive media (including stories, games, and simulations) and share those creations online. Scratch follows in the constructionist tradition – an approach to learning that is grounded in the belief that the most effective learning experiences grow out of the active construction of all types of things, including the construction of computer programs (Kafai & Resnick, 1996; Papert, 1991). The Logo programming environment (developed by Seymour Papert and a team of researchers at MIT in the 1960s) is a significant predecessor to Scratch and has been a major influence on its development. One of the ways in which Scratch has extended the Logo work is in its insistence on the social aspects of learning, most obviously through the creation of the Scratch website.

The Scratch website, launched in May 2007, has become extremely active, with more than 650,000 registered members sharing, discussing, and remixing one another's Scratch projects (Resnick et al., 2009). Each day, members (mostly ages 8 to 16) upload approximately 1500 new Scratch projects to the website – on average, a new project every minute. The collection of projects is incredibly diverse: interactive newsletters, science simulations, virtual tours, animated dance contests, step-by-step tutorials, and many others, all programmed with the Scratch environment and its graphical programming blocks.

Inspired by Papert's Mindstorms, the website draws on the samba school model, where people of all ages come together with shared goals, to support each other's learning and collaborate on endeavors that are more substantial than could be achieved individually (Papert, 1993). More broadly, theories about communities of practice and situated learning have provided ways of thinking about how community settings can support the learning of a practice – by providing learners access to others and opportunities to explore the activities, artifacts, and ideals of the practice (Brown, Collins, & Duguid, 1989; Lave & Wenger, 1991). Accordingly, in addition to enabling people to upload their projects, the site was designed with features typical of community-based content-creation sites like Flickr and YouTube. Members can leave comments on projects, annotate projects with tags, indicate admiration of projects by clicking the *Love It* link, and bookmark others' projects in a list of favorites. Members can also mark other members as friends, create galleries or collections of projects with others, and participate in discussion forums. Each member has a profile page that displays their alias and country, as well as their contributions and interactions – lists of projects, favorites, friends, and galleries.

When asked to describe our work, I usually share a mostly unproblematic narrative about the Scratch website, which we call the *Scratch online community*. This narrative, at least superficially, seems to make appropriate use of the term *community* – over the past three years, hundreds of thousands of young people from around the world have joined and have created more than 1.4 million interactive stories, games, and animations. As designers, we use *community* to express what we desire *for* – and *from* – participants in the technology/environment we have designed, and I less frequently discuss the ways in which the label of *community* might be a mismatch for the activities we observe on the site.

This mismatch of language – the aspirational use of words to create feelings, thoughts, and actions that do not yet (or may never) exist – is obviously not particular to Scratch. It is a side effect of the main activities of designers: bringing possibility to reality, bringing the aspirational to the actual. In this paper, I explore the difference between the aspirational and the actual in the design of online communities, using the Scratch online community as a case study with designer perspectives and participant observations. What are we thinking about when we say *community*? What should the word *community* indicate about the design? How are our intentions supported or undermined by the design? How are the design and our intentions taken up, appropriated, subverted, and transformed by members?

To explore these questions, I conducted interviews with a subset of the MIT Scratch Team known as *metamods*, a sub-group of four people (including myself) that focuses on issues that take place on the Scratch website. The metamods meet weekly to discuss technical extensions (e.g. designing features that limit the impact of spammers) and social extensions (e.g. designing new themed activities for members) to the site, as well as to discuss significant events that have taken place on the site over the week. These interviews and two years of ethnographic observation of the Scratch website were coded to form the basis of the following discussion that explores the gap between the aspirational and the actual in the design of the Scratch online community.

What Does (or Should) the Word “Community” Indicate?

Community is a complicated word – and is made more complex in its translation to the virtual (Bruckman, 2006; Feenberg & Barney, 2004; Rheingold, 2000; Smith & Kollock, 1999). Rather than start with definitions of community from sociology, psychology, or other disciplines, I started the interviews by exploring individual understandings of this word. One metamod described community as something that needs to be explicitly acknowledged by its members:

> People, a place (either virtual or physical), and a sense of belonging by those people who claim to be part of the community. I think people have to be aware, or even explicitly say, that they are part of a community, in order for that community to exist. (Metamod #2)

Another described community as a set of relationships and collaborations, emphasizing the interconnection between people:

> It’s a space where you create relationships and collaborate on different aspects having to do with our lives. To me it is a bunch of people, who are generally in a similar space, who get to have space, time to talk about whatever is relevant and real. They get to perform a lot of different roles for each other – confidante, friend, therapist. (Metamod #3)

Whereas this framed community as activities supporting relationships, another metamod emphasized relationships (or interactions) supporting activities:

> I think about a group of people who share something. What range of things might they be sharing? Something where they either share some activities together, some values together, some goals together. I guess I would typically think it’s a collection of people that have various forms of interactions in support of those shared goals, activities, or values. (Metamod #1)

The metamods’ individual framings of community overlap along (although differently emphasize) several dimensions – individual, interaction, sharing, group – which leads to a working definition of community: *people who intentionally gather together and interact to pursue shared goals and interests*. When asked how this definition of community changes online, metamods talked about how the overall framing and dimensions are the same, but are hindered by communication barriers. It is harder to form intimate interactions, to collaborate as seamlessly, and to express as much as quickly in an environment that lacks non-verbal cues, is asynchronously mediated, and relies primarily on text and images. The nature of online community also provides benefits – connecting with people at a distance and connecting with a much larger, (hopefully) more diverse group of people.

But how do the goals of Scratch and the Scratch website connect with these ideas about community? The overarching constructionist goal of enabling people to have meaningful learning experiences through acts of construction – and, as extended by Scratch, situated in explicitly social environments – can be further articulated into four sub-goals or aspirations: (1) enabling exploration of the *personal*, (2) fostering meaningful *interactions* with others, (3) supporting *diverse* creative expression, and (4) encouraging a sense of group *belonging*. I now discuss each of the aspirations, describing – with vignettes and interview excerpts – the ways in which the *actual* deviates from the *aspirational*.

Aspiration: Enabling Exploration of the Personal / Actual: Pitfalls of the Personal

Computational creation has long been perceived as the domain of a select few, based on gender and epistemological approach (Turkle & Papert, 1990). Scratch aims to broaden the demographic of computational creators, both through technical aspects of the programming environment (e.g. drag-and-drop puzzle-piece-like programming elements that are well-suited to a bricoleur or tinkering approach) and through particular approaches to pedagogical design (e.g. recommending and modeling activities such as collaborative storytelling). Thus, instead of designing activities exclusively framed and focused on implementing certain algorithms or using particular computational concepts such as variables, we encourage young people to work on projects (and suggest that educators design activities) that connect to these concepts via personal interests. For example, creating a project that is an interactive representation of your name or a project about a place you have always wanted to visit or a project about an issue that is important to you. The community should be a place where the personal is explored – but there are pitfalls in exploring the personal online.

Eagle3522 is a 14-year-old girl who creates a Scratch project about the problems that she is currently experiencing in her life. She writes about how she hopes the act of creation will be cathartic, *"I've had a lot on my mind lately, and that's why I made this to kind of try to get it off of my mind."* In the project, she shares fights with her mother, information about her famously incarcerated father, and her depressive thoughts. She receives a mix of supportive and critical comments on her project. She quickly deletes her project, so that it is not visible to the community.

Jacky is a 10-year-old girl who loves ferrets. She creates a series of projects about her favorite animal and the series becomes wildly popular within the Scratch community. She misbehaves and is blocked from accessing the site for a brief period of time. When she returns, she decides that she no longer wants to be defined exclusively by her ferret projects and she deletes all of them, hoping to reimagine her representation and participation.

SonestaMiku is a 14-year-old girl who is active on Scratch, as well as on other media creation sites. She is careful not to share too much personal information on Scratch, but reveals more about her personal interests – and personal life – on these other sites. A friend on Scratch is slighted by SonestaMiku. The friend creates a Scratch project that reveals SonestaMiku's personal information and shares it on the Scratch site. SonestaMiku is furious about this violation and demands that the friend's account be permanently blocked.

All three creators in these vignettes used media creation to express their interests and to share and perform aspects of their identities. But all three were forced to confront key questions. *What types of personal things are suitable for public consumption?* With Eagle3522, the intimate sharing of troubled family life and depressive thoughts became too public. *How does public consumption change what is personal?* With Jacky, the heightened status and visibility of her personal interest was taken on by the community in a way that completely defined her and left her unable to negotiate her self-representation. *Who owns the personal?* With SonestaMiku, the imaginary border around the personal was crossed, and was used against her as a form of sanctioning.

From the metamods perspective, the public-personal is full of contradiction. Exploring the personal is a valuable context for developing as a computational creator, but raises questions about safety – both emotional safety (e.g. the vulnerability of sharing personal problems) and physical safety (e.g. the vulnerability of sharing geographic or contact information). In addition to these questions, there are questions about development. Sharing personal creations publicly in an archive leaves an important trace or record of one's interests and development, but a possibly limiting self-representation. One metamod described this tension, that in our desire to maintain a complete archive we might be losing something valuable:

> We haven't thought about obsolescence, that things need to die. There's some use to having things fade away – and become the humus that we then plant new seeds in. (Metamod #3)

Connecting to and sharing the personal in an online archive is complicated. The personal becomes public and is never forgotten – the desire for *representation* is impinged upon by *visibility*.

Aspiration: Fostering Meaningful Interactions with Others / Actual: Lack of Respect

Creativity is a social process (Sawyer, 2006). It resides not solely in the individual, but through interaction with others. This is a significant reason why the Scratch website is inspired by models of online social networking and community sites, not just models of repositories or archives. Some features of the site were designed explicitly to foster interactions between members. For instance, comments are used to give feedback to others, friending is used to keep track of what others are doing, and forums are used to give and get help with projects. Other features have been unexpectedly appropriated by members for interactions. For instance, the location field is used to give status updates (in a Twitter-inspired fashion) instead of geographic position, and galleries are used as dedicated chat spaces where members flirt. But as with any interactions between people, these interactions are not always positive or respectful.

Suziescape9 is 17-year-old girl who has been creating animations with Scratch for more than two years. She is thoughtful in her creations and her interactions with others. Her work is appreciated by other community members, but positive attention is accompanied by some negative attention. After a particularly thoughtless set of comments, she decides to stop creating projects, which she sees as attracting attention from people she describes as "trolls."

Greg is a teacher who is introducing Scratch to students at his school. He helps them create accounts on the Scratch site and reminds them that shared artifacts are publicly visible. One of his students posts a comment on a stranger's Scratch project, calling it "stupid." The stranger returns the insult, using expletives to describe the student's project. The student is very upset by the exchange. Other teachers find out about it and confront Greg about putting students at risk. Greg feels terrible about what has happened. He sees great benefit in sharing one's work with others online, but the episode has "tarnished" the reputation of the Scratch online community at his school. The students continue to create projects, but no longer share them online.

LiveVictor is a 15-year-old boy who enjoys antagonizing Scratchers whose projects appear on the front page of Scratch. With one Scratcher, he acts as a spammer, leaving numerous gibberish comments on her projects to distract her from other, more productive, comments. With another, he causes panic by pretending to have medical expertise, convincing her that a minor medical issue indicates major future medical problems. After several months of problematic participation, he decides to leave Scratch and makes a project about his departure. In it, he apologizes quite sincerely to those he has harassed and been disrespectful toward. He asks his victims for forgiveness and they accept his apology, and he parts on good terms.

These vignettes illustrate the different reactions members have toward disrespectful interactions. Some are alienated from their creative practices, either entirely like Suziescape9 or partially like Greg and his students. Others accept disrespectful interactions as part of socially-situated creative practices, and learn how to strike back, to just ignore, or – as in the case of LiveVictor's targets – eventually make reparation.

Problematic interactions between members are always a topic of conversation at metamod meetings. One metamod, who deals with complaints from community members and also manages moderators who deal with reports, is particularly focused on ways of responding to disrespectful interactions. He views one of his most significant contributions as creating a safe space where creative practices can flourish:

> I've seen that one of the important things I do is to create safety, in a sense. I think that community is an emergent phenomenon – it's something that creates itself. You give people the right kind of context, and they start to create new ideas, and share them, and grow them. But that starts to break down when something becomes unsafe, when people feel threatened or they feel attacked. They contract into this posture that's very defensive, basically, and there's less of the sharing, growth, and interesting phenomena that I like to see happen. So, I do that by removing threats or really disrespectful content. (Metamod #3)

Another metamod acknowledged that there was more that could be done in a proactive fashion to encourage community-appropriate interactions – rather than re-establishing a sense of safety *after* a negative interaction occurs between members:

> There are people who don't show, in my mind, appropriate respect for others in the community. It's not that it's set up in a way that there's necessarily as high a level of respect among people in the community that I would think would be ideal. And one could imagine certain communities do more to support that level of respect. Everything from certain initiation routines and rituals that help people get a better understanding of the values of the community, which I don't think there's a lot of in the Scratch community. (Metamod #1)

For example, it is currently quite easy to join the community. An applicant fills out a simple form, specifying a username, password, birth date, email, gender, and country. There is no other text on the page that indicates what the individual can do with the account or what expectations exist regarding participation. Those expectations are shared in the Scratch community Terms of Use, which are accessible as a link at the bottom of the page, wedged between "Download | Donate | Privacy Policy" and "Copyright Policy | Contact Us." The Terms of Use (http://info.scratch.mit.edu/Terms_of_use) are expressed as six expectations: (1) Be respectful, (2) Offer constructive comments, (3) Give credit, (4) Be honest, (5) Keep personal information private, and (6) Help keep the site friendly. Important (perhaps even commonsense) ideas, but members are usually only exposed to the terms after the terms have already been violated.

Routines and rituals to clarify expectations about participation would help new members better understand this new environment – and clear communication of expectations seems even more important in an online space where communication is already challenging (as described earlier). Many come to online spaces with an exaggerated sense of invincibility afforded by anonymity, often behaving in disrespectful ways that they would not in real life. Thus, feelings of *safety* within the existing community are impinged upon by the *anonymity* of the not-yet-enculturated individual.

Aspiration: Supporting Diverse Creative Expression / Actual: Conflicts in Interests

When describing the aspirations of Scratch, we frequently use the metaphor of a room (Resnick & Silverman, 2005). We want Scratch to have a *low floor* – it should be very easy for anyone to get started creating things with Scratch. We also want Scratch to have a *high ceiling* – people should be able to create projects of arbitrary complexity and not be limited by the easy beginnings. Finally, we want Scratch to have *wide walls* – people should be able to create a wide range of projects, from adventure stories to role-playing games to science simulations, not just a single type or genre of project. While we see members experimenting with different genres of projects, many members tend to focus on a particular genre, which varies most obviously according to gender and age.

Dvora160 is 14-year-old boy who is opposed to (what he perceives as) the superficiality of animation or other aesthetically-focused projects. He and others are adamant about the primacy of programming in Scratch and argue that technically-sophisticated projects should be valued most highly in the community. He creates a protest project, encouraging people to boycott all animation projects and calling for a return to a programming focus.

Ashleigh_Jaguar is a 16-year-old girl who is a passionate animal rights activist. She creates a Scratch project that includes graphic photos of animals being mutilated in product and drug testing. She posts the project and encourages others to support her in the animal rights cause. A young Scratcher views the graphic content and is scared. His parents contact the Scratch Team, which then removes the content, explaining that the content was too intense for the broad Scratcher age range. She defiantly re-posts the project, stating that the truth of animal testing is too important for the Scratch Team to ignore or suppress.

Zaw is a 15-year-old boy who disdains Scratch's core values of personal expression and computational creation. He wants to destroy the Scratch community. He decides to recruit members to join an army that will systematically drive people away from Scratch. Members of the army are encouraged to post negative comments on people's projects until the creators leave Scratch. His negative attitude and scare tactics incite members who care about the Scratch community and its core values. Zaw is eventually banned from the community, but his army's presence lingers through old comments on projects, occasionally causing panic in those who fear the return of Zaw.

Members are passionate about their genres, interests, and goals. But sometimes their interests are seen as being at odds with others' interests and modes of expressing those interests. As the size of the community increases, these conflicts also increase. There are more people with similar interests with whom a member can align her or himself. There are also more people with perceived opposing interests that a member can align against, as with Dvora160 versus the animators. These conflicts of genre preferences and issues of age-appropriateness (as with Ashleigh_Jaguar) can be handled through moderator mediation and respectful interactions between involved parties. The case of Zaw is somewhat unique, in that he fundamentally rejects the overarching goals of Scratch. One metamod described the difference between these two different conflicts:

> The community is not homogeneous. There are shared values and shared goals, but it doesn't mean that everyone has exactly the same goals or exactly the same values or exactly the same interests. Ideally, I'd like to see the values as more consistent and more homogeneous. Whereas the interests would vary and there would be greater diversity, there should be more openness, with many different pathways for interests or styles of work. (Metamod #1)

But this assumes that one is even able to identify interests that are shared (or not) by others. All of the metamods described the difficulty that the size of the Scratch website presents to members, both new and old. One metamod described how he saw the inability to find one's interests reflected in the archive as contributing to negative feelings:

> The challenge is that you can point to this place and say who we are, but it's really hard to look at it from outside. In Scratch, it's really hard because you have to either search or – there aren't the right affordances. I feel that's a problem with most large archives, online archives. There aren't the right tools yet to see everything and organize it and zoom in and zoom out. I think it makes people feel overwhelmed and makes them feel less powerful. (Metamod #2)

In either case, the value of *diversity* is impinged upon by either a desire to protect one's sense of – or lack of inclusion in – *identification* with others.

Aspiration: Encouraging a Sense of Group Belonging / Actual: Seeking Attention

Collaboration enables people to construct ideas and artifacts greater than they could achieve independently (Koschmann, 1996; Rogoff & Lave, 1984). There are numerous capacities that contribute to effective collaboration, including communication and perspective-taking capacities. Collaboration occurs in different configurations and with different aims. In Scratch, we see people collaborate in pairs, small groups, and large groups – with both top-down and more diffuse management styles. These configurations support all aspects of an iterative design process – from imagining projects, through creation, playing, sharing, and reflection (Resnick, 2007). In a fundamental way, the entire Scratch community is a large collaborative experience, with people exchanging ideas to contribute to a collective progression as computational creators, expressing themselves creatively through media artifacts. But this general sense of belonging to the larger social sphere is sometimes undermined by the position and progression of the individual – the ways in which the *we* is undermined by the *me*.

Puppet is an 11-year-old boy who has grown savvy (and somewhat cynical) about tactics that Scratch members use to become famous on the Scratch site. He creates a project about how to become famous, which he describes as partly ironic and partly accurate. He recommends spamming members' projects, creating projects that declare a dramatic and abrupt departure, and establishing memes that will be broadly appealing to others and popular in the community. Other Scratchers don't notice – or choose to ignore – the ironic intention of the project and post comments about which of the approaches (and suggestions for other approaches) have worked for them.

The_Destroyer is a 12-year-old boy who threatens to take down the Scratch community server and members' computers with a virus that he has written. He writes threatening messages to others in the community on the forums and encourages them to acknowledge his amazing powers. SNYgames is a Scratcher who rushes to the community's defense, and courageously announces that he will confront and defeat The_Destroyer. The drama between SNYgames and The_Destroyer escalates, causing much anxiety in other Scratch members. A moderator defuses the situation by publicly exposing SNYgames as The_Destroyer, and kindly asks that SNYgames cease the dramatic impersonation.

HiddenComet is a 14-year-old girl who shares her experiences of having diabetes with the Scratch community. There is an outpouring of support and concern for her in the face of this life-threatening condition. Another member of the community, Chattydeign, shortly announces that she too has diabetes and is on the verge of death. It is revealed, however, that Chattydeign is deceiving community members to gain attention. Members create projects about the deception and harshly criticize Chattydeign for attempting to falsely gain sympathy and status.

Attention is important – Scratch community members want to be recognized for their contributions. A metamod described a lack of attention as obviously demotivating:

> When you don't get any feedback for what you create, that undermines the sense of community because you feel like nobody is listening to you. There's no difference between you being by yourself, doing what you were doing before and doing it now in this space. (Metamod #2)

He went on to explain how the size of the community can contribute to further demotivation, if one is unable to *ever* imagine oneself as making a major contribution:

> The fact that we have libraries with tons of amazing books makes me less likely to feel that I'm going to be a great writer. There are all these amazing people. I'm just going to write a paper – I'm never going to be the next Shakespeare or someone like that. I feel that, in the same way, it could detract from people feeling like they could make a big contribution. As the site grows, you feel smaller. (Metamod #2)

But a need for attention is fundamentally in conflict with a sense of community. Attention is not equally distributed across members, and there is only so much attention available. These conditions lead to members becoming problematically tactical with the community's attention, as demonstrated by the vignettes about Puppet, The_Destroyer/SNYgames, and Chattydeign. The metamods group regularly discusses (and designs) ways of productively responding to the desire for attention, in ways that benefit both the individual and the community. From featuring projects with less visibility to promoting the value of collaboration, our responses are centered on cultivating awareness of others – a me-we, self-social progression:

> They're at this stage, "It's me! What can I get?" and that's the appropriate moral development that they should be at. But at the same time, I want to provide some sort of scaffolded question that leads them to think – "What is the entire effect on the community?" – so they can acknowledge that what they're participating in is a real thing. It's often moving people, helping moving people from a totally self-focused view to a slightly larger view. Whatever slightly larger means to them. (Metamod #3)

But is it a Community?

These four gaps between the aspirational and the actual demonstrate a common problem: it is increasingly difficult to connect with others – and multiple factors amplify this difficulty. The size of the site is a problem, not simply because there are many people participating, but because there is a lack of satisfying ways for representing or accessing that large number of people. The public, anonymous, and reduced expression of interactions on the site is another problem that continually impacts people's ability to form intimate relationships with one another. The natural tension between the self and the social challenges connections between people, from desires for fame to fears of loneliness. Finally, there is the difficulty of *exclusively* valuing problem-free connections or relationships with others. While some may be alienated by conflict, it can simultaneously unite others. Drama, nemeses, and hardship can all serve to connect people, even if at the expense of others.

Despite these gaps between the aspirational and the actual, is the Scratch website a community? This is not a preoccupation with the word *community*, but rather with whether or not we (as metamods, as the Scratch Team) are satisfied with the ways in which the website represents our aspirations. Is it a community in the way we desire it to be a community? I asked each of the metamods whether they felt the Scratch site is a community. One described how, although he saw many ways in which interactions between members could be improved, he saw relationships being formed as an indicator of community:

> Yes, I'd say that Scratch is a community. And the evidence of this has to do with the cultivation of long-term relationships, which does occur a lot on the Scratch website and the expression of – what's the word? for brotherly love? agape – that "you and I we're sort of similar." There's that kind of acknowledgement. (Metamod #3)

Although generally not concerned with the use of the word *community*, another metamod adopted a similar litmus-test approach to the Scratch-website-as-community question. His focus was on activity and values:

> I tend to think of it as a community. I don't focus so much on what the definition of community should be. But I see the Scratch community as both sharing some activities – of creating these particular types of artifacts with this particular tool. But I'd also like to see it as a community that shares certain values – the values of personal expression, collaboration. (Metamod #1)

Consistent with his initial definition of community, one metamod talked about the subjective nature of the Scratch community, that it "is more of a personal decision." When I asked directly whether it is a community *for him*, he described how our differentiated status as designers complicates the relationship:

> For me? No. I guess it's hard. Because the position we have – it's like the mayor of a city, who I guess is part of the community, but you have a different position. I don't feel like I'm part of the kids. It seems like there is some distinction there, by the way the kids relate to the Scratch Team members. They know they are different from us. For me, I do recognize it as a community, and I feel part of it insofar as my role as Scratch Team member. My experiences are probably different from those who are "real users" of the community. (Metamod #2)

Our role is obviously complicated. We are designers – participants with intention and power. But what does this difference imply about the relationship between the designers and the non-designers? All three metamods adopted different perspectives on the significance or importance of the designers in defining and driving community processes. One viewed active participation and community cultivation by the Scratch Team as important, but acknowledged that competing interests between designers and other members may result in tension:

> There can sometimes be a tension between being a designer – of having certain types of goals and values that you see as important for the community – but then also wanting to support the desires and goals of the community members. Sometimes there can be a conflict between what some people in the community see as the values and goals of the community. That can be a challenge. (Metamod #1)

Another described definition and participation as a joint responsibility of designers and other members. Members should feel a responsibility for making the community, but that the making is overseen by the designers:

> It's more that this community is yours, and if you want to do something different, I encourage you to do that. As long as it's positive and helpful. (Metamod #3)

Finally, one described authentic participation as emerging from ownership, and wished that the members could be given more autonomy over *their* community:

> I feel that one thing we could do to make it more of a community is to have the ability for people to decide what are the things where they can participate for the benefit of community

> and let them select those things and let them do those things. *So you see the Scratch Team as external to the community processes?* Yes. (Metamod #2)

Mind the Gap: Reflections and Suggestions

This work began with a question about using the word *community* to describe the Scratch website. The process of coding ethnographic observations of the community and interviews with the community's stewards has helped me think about this question – and has provoked new questions. Although I knew at the beginning that *community* is a complex and multifaceted word, having the opportunity to unpack some of that complexity in conversation with other metamods was quite valuable. While we share general conceptions of what the word *community* means, it is obvious that we value different aspects of what *community* represents.

The newly-provoked questions do not stem from a definitional preoccupation. They emerge from an acknowledgement that design of technological infrastructure is insufficient to support the aspirational – the goals of the Scratch project and of constructionism more generally. It is not enough to stand by and simply observe what happens with (or to) the site, to assume that our existing design is sufficient in and of itself, and that members will not take up and reimagine the space in conflicting ways. de Certeau (1984) cautioned against making this assumption:

> It is nonetheless implicit in the "producers'" claim to "inform" the population, that is, to "give form" to social practices. … To assume that is to misunderstand the act of "consumption." This misunderstanding assumes that "assimilating" necessarily means "becoming similar to" what one absorbs, and not "making something similar" to what one is, making it one's own, appropriating or reappropriating it. (p. 166)

We need to be actively contributing to, participating in, and negotiating the interactions and activities on the site to support our goals. We must constantly struggle to recognize and to mind the gap between the aspirational and the actual in the online community. Yes – community is subjectively defined. Yes – community is subjectively experienced. But that does not mean that we should refrain from influencing its development. As designers, we should continually ask if it is moving toward the actual to which we aspire.

References

Brown, J.S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. Educational Researcher, 18(1), 32-42.

Bruckman, A. (2006). Learning in online communities. In K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (pp. 461-472). Cambridge: Cambridge University Press.

de Certeau, M. (1984). *The practice of everyday life*. Berkeley: University of California Press.

Feenberg, A., & Barney, D. (Eds.) (2004). *Community in the digital age: Philosophy and practice*. Oxford: Rowman & Littlefield.

Kafai, Y.B., & Resnick, M. (Eds.) (1996). *Constructionism in practice: Designing, thinking, and learning in a digital world*. New Jersey: Lawrence Erlbaum.

Koschmann, T. (1996). *CSCL: Theory and practice of an emerging paradigm.* New Jersey: Lawrence Erlbaum.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.

Papert, S. (1991). Situating constructionism. In S. Papert & I. Harel (Eds.), *Constructionism*. Cambridge, MA: MIT Press.

Papert, S. (1993). *Mindstorms: Children, computers, and powerful ideas* (2nd ed.). New York: Basic Books.

Rheingold, H. (2000). *The virtual community: Homesteading on the electronic frontier*. Cambridge: MIT Press.

Resnick, M. (2007). All I really need to know (about creative thinking) I learned (by studying how children learn) in kindergarten. *Proceedings of ACM Creativity & Cognition conference, Washington DC, June 2007*.

Resnick, M., Maloney, J., Monroy-Hernandez, A., Rusk, N., Eastmond, E., Brennan, K., Millner, A., Rosenbaum, E., Silver, J., Silverman, B., & Kafai, Y. (2009). Scratch: Programming for all. *Communications of the ACM, 52*(11), 60-67.

Resnick, M., & Silverman, B. (2005). Some reflections on designing construction kits for kids. *Proceedings of IDC conference, Boulder, CO, June 2005*.

Rogoff, B., & Lave, J. (1984). *Everyday cognition: its development in social context*. Cambridge: Harvard University Press.

Sawyer, K. (2006). *Explaining creativity: The science of human innovation*. Oxford: Oxford University Press.

Smith, M.A., & Kollock, P. (Eds.) (1999). *Communities in cyberspace*. New York: Routledge.

Turkle, S., & Papert, S. (1990). Epistemological pluralism: Styles and voices within the computer culture. *Signs, 16*(1), 128-157.

Supporting Collaborative Learning in Recitation Sections Using an Ambient Awareness Tool

Hamed S. Alavi, Pierre Dillenbourg, Swiss Federal Institute of Technology, 1015 Lausanne, Switzerland
Email: Hamed.alavi@epfl.ch, pierre.dillenbourg@epfl.ch

Abstract: We study the effect of an ambient awareness tool, named *Lantern*, on the quantity and structure of collaboration in recitation sections (i.e. when students work in small teams on the exercise sets with the help of teaching assistants). Lantern is an interactive lamp that with colors, intensity of light and blinking represents a piece of information. In a Lantern-equipped recitation section, every team is provided with one Lantern which shows the current status of that team: the exercise they are working on, if they called the teaching assistants for help, and since when. The results show that (1) Lantern increases intra-team collaboration while teams are waiting for teaching assistants and (2) it increases the diversity of inter-team communication, that is, each team communicates with a larger number of other teams, and there are fewer teams who never communicate with others.

Introduction

In university teaching, recitation sections are sessions in which students solve pre-assigned problems while one or more teaching assistants (TA) provide support by answering questions. These sessions supplement lectures by providing an informal learning atmosphere that allows collaboration among students and supports it by supervision from the TAs. Consequently, the efficiency of a recitation section highly depends on the level and structure of collaboration among students as well as the effectiveness of TA supervision. In order to get insight on the efficiency of the recitation sections, we observed and videotaped 12 recitation sessions in our school. The analysis on the recorded data showed some problems. For example, (1) students spend a considerable amount of time and cognitive effort to catch the attention of TAs when they need help, (2) they usually have a limited circle of friends with whom they collaborate, and (3) in many cases, when choosing which team to assist, the TAs fail to notice or respond to more pressing requests.

We believe that many of the shortcomings are due to a lack of awareness information both on the part of the TA and among the students themselves. We thus proposed an awareness tool, called *Lantern*, which gives information on the status of students: who is working on what exercise, who has asked for help and since when. In our previous work we have shown that Lantern can improve the efficiency of interaction between students and TAs (Alavi, Dillenbourg, & Kaplan, 2009). We were also interested to find out whether informing students about the status of the others would encourage a higher level and more complex structure of collaboration. This paper focuses on answering the question of how does Lantern affect collaboration in recitation sections? For instance, from our previous study we know that, when using Lantern, students are much more productive while waiting for the TA. An interesting question is whether Lantern also increases collaboration in this waiting time.

Our comparative analysis on six Lantern-equipped and six regular recitation sections shows that Lantern increases the intra-team collaboration when students are waiting for TA. Moreover, Lantern changes the structure of inter-team communication, such that each team communicate with a larger number of other teams and that there are fewer teams who never communicate with others. Please note that we intentionally avoid referring to the interactions taking place among groups as collaboration. On the other hand, the joint effort within a team to solve exercises validates the criteria of collaborative learning (Dillenbourg, 1999).

The reminder of this paper is organized as follows. First we explain the recitation sections as the context of this work. The next section describes our tool Lantern. Then we explain our past experiment along with a summary of its results followed by the new research questions. We answer the questions, through an experiment described in the next section as well as the analyses on the collected data. We summarize the relevant research and finally draw some conclusion.

Context: Recitation Section

In general, the recitation section is a complement to the lecture session as a university pedagogical practice. Students work on their assignments, individually or in small groups. Depending on the demand of the class, a number of teaching assistants are present to give help, hints or in some cases public explanation. When they need help, students raise their hand and wait for the TA.

In particular, the recitation sections we studied shared the following properties:

- Attending the class was not mandatory for the students.

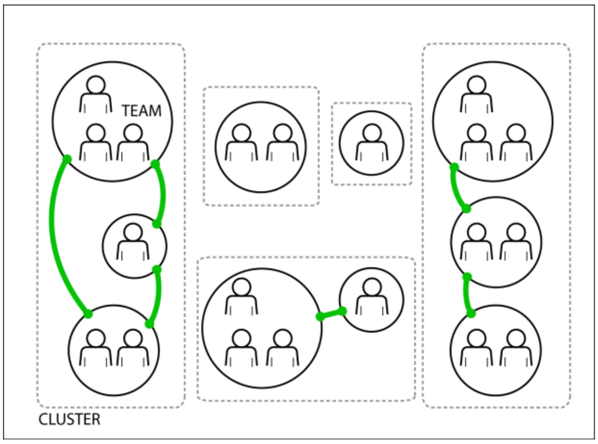


Figure 1. Structure of Interactions in Recitation Sections.

- 20-35 students attended each session, working individually or in small groups of at most 6.
- Figure 1 shows the structure of collaboration in a recitation room. Teams sat closely and went through the exercises together. They had limited interactions with some of their neighbors. In a session, there are 3-5 completely disconnected *clusters* of teams.
- 1-3 TAs ran each session.
- Each session lasted about 2 hours, while students were allowed to leave earlier.
- 3-8 exercises were assigned for each session.
- The exercises were usually theoretical in that they only involved pen and paper.

Tool: Lantern

Lantern, shown in Figure 2, is a small and portable lamp which consists of five pairs of Light-Emitting Diodes (LEDs) installed in a column and covered by a blurry plastic cylinder. A microprocessor controls the LEDs. Each team is provided with a Lantern, which makes use of a very simple visual grammar to show the status of that team:

- **Color:** The color of the team's Lantern indicates the exercise they are currently working on.
- **Intensity of light:** It grows up with time on the five floors of LEDS and specifies the time that has been spent on the current exercise.
- **Blinking:** It indicates that the team is calling the TA for help.
- **Frequency of blinking:** The rate of blinking increases with time, showing for how long the team has been waiting for the TA.

Users can interact with Lantern by turning it to choose an exercise and by pressing on it to call for help. Each lantern records all user interactions, which can be downloaded through a USB connection for analysis.

Figure 2. Lantern.

Past Observations

We started by conducting a field study in which we watched and videotaped three recitation classes for four consecutive weeks. The analyses on this observation led to the idea of Lantern. To evaluate the first prototype of Lantern, we set up an experiment in which Lantern was used in two courses through four weeks of recitation. The main result was that the *productivity* of students, *while waiting* for the TA, improved remarkably. The details of the field study and the first experiment along with the analyses and results are reported in our previous work (Alavi et al., 2009). In the following, we recall the concept of *while-waiting productivity* and use it to explain our new research questions.

While-waiting Productivity

According to our observations, in the control condition, when teams have to wait for the TA, they spend a considerable amount of time chasing the TA, i.e. trying to catch her attention. We defined while-waiting productivity as the fraction of the waiting time that is not spent on the chasing. We use this parameter as an indicator of the teams' efficiency during this waiting time. In the first experiment, students had 94% while-waiting productivity when they used Lantern and 38% when they did not (Alavi et al., 2009).

Note that the waiting time can be considerably long, (depending on the number of demands and the number of TAs). Table 1 shows how long, in average, a team had to wait for the TA over a session, in the field study and the first experiment. Furthermore, the performance of students during the waiting time is of special interest, since this is usually a high value period when students are challenged with and focused on the exercises.

Table 1: Waiting time in actual recitation sections.

	Field study	First experiment (with Lantern)
Avg waiting time (min)	21	16
Max waiting time (min)	36	27
Avg session length (min)	98	87

Research Questions

A while-waiting productivity of 94% means that students use 6% of their waiting time chasing the TA. An interesting question is what do students do during this 94%? Specifically, do they collaborate and what is the influence of Lantern on that?

We believed that, Lantern can encourage both intra-team collaboration and inter-team communication, especially when students are waiting for TA. If members of a team need help while trying to solve a problem together, they can press on Lantern and continue the collaboration, whereas without Lantern, they usually stop the collaboration and start chasing the TA to get her attention. On the other hand, if a team waiting for help on an exercise realizes that their neighbors have already solved that exercise may ask from that team, which triggers inter-team communication. We thus, posed the following questions:

- *Q1.* Does Lantern increase intra-team collaboration while the team is waiting for TA?
- *Q2.* Does Lantern increase intra-team collaboration while the team is NOT waiting for TA?
- *Q3.* Does Lantern increase inter-team communication while the teams are waiting for TA?
- *Q4.* Does Lantern increase inter-team communication while the teams are NOT waiting for TA?

We were also interested to evaluate the effect of Lantern on the structure of inter-team communication in the classroom level (Figure 1):

- *Q5.* How does Lantern affect the structure of communication among teams?

To answer the above questions, we designed an experiment which is explained in the next section.

Recent Experiment

Third-year students of computer and telecommunication sciences are observed during six regular and six Lantern-equipped recitation sections. Table 2 shows the basic parameters of the sections.

Table 2: Observed sessions in the control and Lantern conditions.

Condition	# sessions	# students	#teams	team size
Control	6	23-34	12-14	1.7-2.4
Lantern	6	20-29	10-13	1.8-2.2

Data Collection

One observer (the first author of this paper) tried to note every interesting event, especially the collaboration among students which are mostly recognizable as verbal interactions and certain body postures. Considering the small size of the classrooms, it was not a problem to distinguish the topic of the conversations (relevant or irrelevant to the course material). More precisely, the observer, with a one-minute precision, registered whether each student is (1) working individually, (2) collaborating with a teammate, (3) communicating with another team, or (4) not engaged in the exercise set (e.g. Table 3).

Table 3: The work status of the students.

time	Student 1	student2	student3	…	student34
min1	4	4	4		4
min2	1	1	1		4
min3	1	2	2		1
…					
min120	4	4	3		3

In addition, from the data logged by Lantern we know about the status of each team: when they start and finish an exercise, when they call for, and receive help. Altogether, the following data is collected from each session:

1. When a student starts collaborating with another member of her team and when they finish the collaboration.
2. When a student starts communicating with another team and when they finish the communication.
3. When a team calls for help, when the TA arrives to give help, and when the TA leaves the team.

The collected data also can be represented by the status of students, as visualized in Figure 3. For example in this figure, the student starts collaborating with her teammates at time *t1*. At time *t2* her team calls for help and starts waiting for the TA who arrives at *t4*. Meanwhile she stops the collaboration at *t3*. The TA leaves the team at *t5* and she goes back to the problem solving mode.

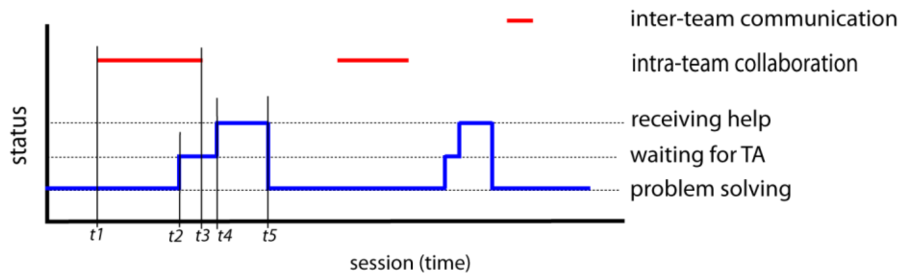


Figure 3. Status of a Student during a Recitation Section.

At the end of the experiment, the students and the teaching assistants completed a questionnaire about their experience with Lantern. We will refer to the responses when we try to explain some of our findings in the next sections.

Analysis

In this section, we analyze the influence of Lantern on intra-team collaboration and inter-team communication. We consider six Lantern-equipped recitation sections and compare them to six regular sessions as the control condition.

Intra-team Collaboration

We define four parameters for each session:

- Col_C^W, Col_L^W is the percentage of the ***W***aiting time each student spent on intra-team collaboration, averaged in the session, in the ***C***ontrol and ***L***antern conditions respectively.
- Col_C^P, Col_L^P is the percentage of the ***P***roblem solving time each student spent on intra-team collaboration, averaged in the session, in the ***C***ontrol and ***L***antern conditions respectively.

Note that, in our analysis, we exclude the time intervals when the team is receiving help (like *t4* to *t5* in Figure 3). This is because interaction among students is highly influenced by the presence of the TA, while receiving help.

We start by comparing Col_L^W against Col_C^W. For the case of collaboration that begins before the waiting period and lasts through it (for example, the first collaboration in Figure 3), the second part (*t2* to *t3*) counts in the Col^W values. An unpaired t-test shows a significant improvement with Lantern ($m_{lw} = 42, m_{cw} = 11.17, t[10] = 10.97, p < 0.0001$): positive answer to question Q1.

One may explain this improvement as the direct effect of the improvement in while-waiting productivity. In order to test that, we modify the measures as follows:

$$\overline{Col}_C^W = \frac{Col_C^W}{WWP_C} \text{ and } \overline{Col}_L^W = \frac{Col_L^W}{WWP_L}$$

In which WWP is the average while-waiting productivity of the session. An unpaired t-test shows that $\overline{Col}_L^W$ is still significantly higher than $\overline{Col}_C^W$ ($m_{LW} = 47.1, m_{CW} = 24.2, t[10] = 5.8, p < 0.001$).

We found no significant difference between Col_L^P and Col_C^P ($m_{LP} = 33.8, m_{CP} = 32.6, t[10] = 0.40, p > 0.1$), meaning that Lantern has no significant effect on the intra-team collaboration when the team is not waiting: negative answer to question Q2.

We were also interested to know if students collaborate more when they are waiting than when they are in problem solving mode, in either of the conditions. To do that we compare $\overline{Col}_C^W$ to Col_C^P, and $\overline{Col}_L^W$ to Col_L^P using a paired t-test. Interestingly enough, we found out that, in the control condition, students collaborate less than usual when they are waiting ($m_{CW} = 24.2, m_{CP} = 32.6, t[10] = -2.62, p < 0.05$), whereas in the Lantern condition students collaborate more when waiting ($m_{LW} = 47.1, m_{LP} = 33.8, t[10] = 8.19, p < 0.001$).

Figure 4 summarizes the result of the above comparisons.

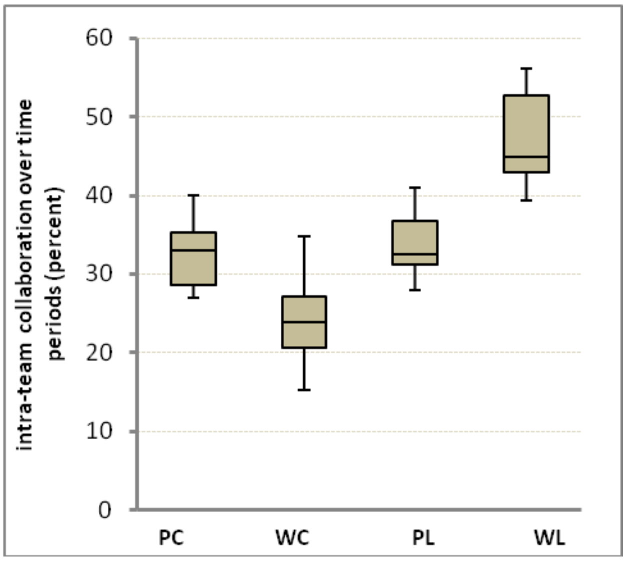


Figure 4. Intra-team Collaboration across the Conditions.
(PC: $\overline{Col}_C^P$, WC: $\overline{Col}_C^W$, PL: Col_L^P, WL: Col_L^W)

Inter-team Communication

This section analyzes the effect of Lantern on inter-team communication in terms of the duration of the interactions as well as their structure.

Duration

Our analyses show that Lantern does not significantly increase the duration of inter-team communication either while the communicating teams are waiting for the TA or while they are doing problem solving: negative answer to questions Q3 and Q4.

Structure

In order to quantify the effect of Lantern on the structure of inter-team communication, we define the following interrelated parameters, for a session:

- *Diversity of Communication:* is the total number of pairs of teams which communicate at least once in the session. In Figure 5 each link between two teams indicates an inter-team communication. In this figure the Diversity of Communication is 6.

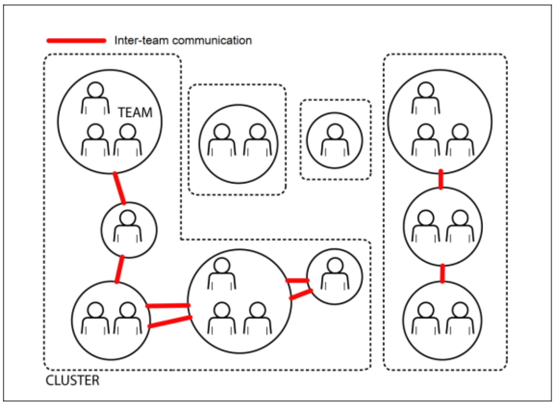


Figure 5. Inter-team Communication in Recitation Sections.

- *Number of Communicating Teams:* is the total number of teams who were involved in inter-team communication at least once. In Figure 5 there are 8 Communicating Teams.
- *Number of Clusters:* a Cluster consists of a set of teams such that, each team in the Cluster communicates with at least one other team of the same Cluster, and there exist no communication between two teams of two different Clusters. In Figure 5 the number of Clusters is 4. In an interpretation that considers teams as sources of knowledge, a recitation section with fewer clusters provides a better platform for the flow of knowledge.

Each parameter is averaged over the number of teams in the sessions and analyzed using an unpaired t-test. The results show significant increase in all the parameters. As a sample, the result of the t-test on the Diversity of Communication is ($m_L = 0.39, m_C = 0.26, t[10] = 3.52, p < 0.01$), summarized in Figure 6.

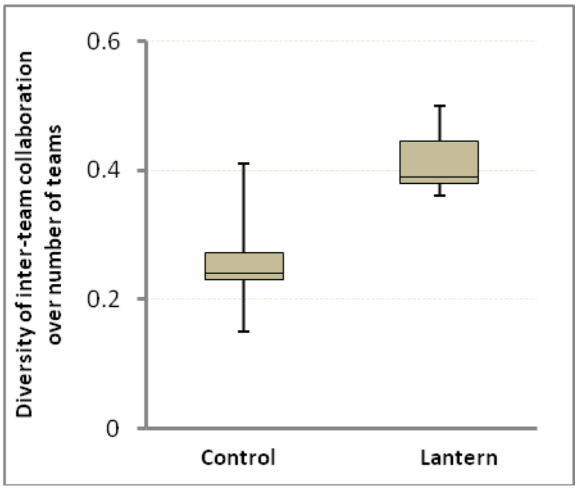


Figure 6. Diversity of Communication across Condition.

Summary of Results and Discussion

Lantern increased intra-team collaboration while the team is waiting for the TA. This can be explained by the combination of two facts: (1) according to the questionnaire, before pressing on Lantern the team members agree that they need help, which itself initiates a discussion, and (2) this discussion can last through the waiting time, since no effort to catch the TAs' attention is needed. The quantity of the improvement, comparing to the total time one student interact with her teammates in the control condition in average, is 22.1%. Moreover, it has been shown that, with Lantern, students do more intra-team collaboration when they are waiting for the TA than when they are doing problem solving. *We conclude that, when students are waiting for the TA there is a high potential for collaboration, which is lost in chasing the TA when Lantern is not used. As a consequence, the waiting time does not always need to be shortened: in some cases, using an awareness tool rather than adding new teaching assistants would maintain longer waiting times that become fertile grounds for collaboration to take place.*

Lantern did not increase the duration of inter-team communication. However, it has been shown that Lantern has effect on the structure of inter-team communication. More precisely, when Lantern is used each team communicates with a larger number of other teams (higher Diversity), (2) there are fewer teams who never communicate (more Communicating Teams), and (3) the knowledge of one team can be spread over a larger part of the class (fewer Clusters). We explain this set of effects as the direct consequence of the main objective of Lantern: adding to the knowledge of students about other teams. The more students know about a specific team the more they are likely to interact with that team.

Related Work

In this section, we position our work among the relevant research from three different fields: (1) Computer Supported Collaborative Learning (CSCL) research on tools for regulating teams' interactions, (2) the Computer Supported Cooperative Work (CSCW) research on awareness tools and (3) the work of ambient interface and reality-based interaction in Human-Computer Interaction (HCI).

In CSCL, Soller et al. (2005) provided a framework that categorizes collaborative learning supporting systems into three classes: (1) mirroring systems, which display raw indicators to collaborators (2) metacognitive tools, which monitor the interactions, process the collected data and represent the state of interaction via a set of high-level indicators (3) coaching systems, which offer advice based on an interpretation of those indicators. We make use of this framework to compare our work against the others. Lantern fits in the first category as they mirror the state of student groups to the groups themselves and to the TAs without any pre-processing.

Chen (2006) designed a tool, called Assistant, which monitors the collaboration, visualizes the processed data and provides advice to the teacher. It can also learn from teacher's feedback to improve its performance. Assistant should be put in the third category of Soller's framework (coaching systems). Moreover, Assistant is tailored for the context of distance collaborative learning, while Lantern is designed for co-present settings.

Avouris et al. (2004) developed a collaboration environment called Synergo, for collocated and distance learning. Synergo monitors the activity, makes analyses and visualizes quantitative parameters like density of interaction, symmetry of partner's activity etc. It also provides teachers with useful information to manage the interactions that occur in the classroom. Synergo fits in the middle category (metacognitive tools).

Our work is also different than Chen's and Avouris' in terms of the level of interaction it considers. While Assistant and Synergo are mostly centered on interactions within one group, we are looking at the higher level, i.e. interaction between several groups and TAs as well as the interactions among groups.

In CSCW, there have been many efforts aiming for providing awareness information. They can be categorized in three dimensions:

1. The temporal nature, i.e, if the information is given and used at the same time and context as it is collected (synchronous (Beaudouin-Lafon & Karsenty, 1992; Ishii et al. 1994; Shen & Sun, 2002)) or not (asynchronous (Manohar & Prakash, 1994))
2. The type of provided awareness information which can be about workspace (Fitzpatrick et al., 1999; Fitzpatrick et al., 2002; Fuchs et al., 1995; Ishii et al., 1994) or presence (Shen & Sun, 2002; Tang & Rura, 1994), and activities of participants in the cooperative work.
3. The context, including the users, task and the structure of interaction among participants (e.g. conferencing (Shen & Sun, 2002), and distance learning).

Lantern gives *real-time* information, on students' activity in a *collocated* collaborative *learning context*.

Finally, in HCI, the seminal idea of ambient interface is to extend classical user interfaces (display, keyboard, mouse) to the whole environment. In contrast to the works described above, the primary concern of ambient display applications is the subtle embedding of information in our surroundings, while capturing and processing information is of a minor concern. The effectiveness of ambient interfaces for providing awareness information has been shown in many cases (Heiner et al. 1999; Pederson & Sokoler, 1997). In line with those works, we have proposed Lantern as a tool that embeds awareness information into the recitation classrooms, and evaluated its capability to support collaborative learning.

Conclusion

This paper analyzes the influence of Lantern on the collaboration in recitation sections. We show that, Lantern increases the intra-team collaboration taking place while the team is waiting for the TA. Lantern also has effect on the structure of inter-team communication in the recitation sections in such a way that, each team communicates with a larger number of other teams and that there are fewer teams who never communicate with others.

References

Alavi, H. S., Dillenbourg, P., and Kaplan, F. (2009). Distributed Awareness for Class Orchestration. In *Proceedings of the 4th European Conference on Technology Enhanced Learning: Learning in the Synergy of Multiple Disciplines* (Nice, France, September 29 - October 02, 2009). U. Cress, V. Dimitrova, and M. Specht, Eds. Lecture Notes In Computer Science, vol. 5794. Springer-Verlag, Berlin, Heidelberg, 211-225.

Avouris, N., Margaritis M., and Komis V. (2004). Modelling interaction during small-group synchronous problem solving activities: The Synergo approach, *2nd Int. Workshop on Designing Computational Models of Collaborative Learning Interaction,* ITS 2004, 7th Conf. on Intelligent Tutoring Systems, Maceio, Brazil, 2004, 13–18.

Beaudouin-Lafon, M. and Karsenty, A. (1992). Transparency and awareness in a real-time groupware system. In *Proceedings of the 5th Annual ACM Symposium on User interface Software and Technology* (Monteray, California, United States, November 15 - 18, 1992). UIST '92. ACM, New York, NY, 171-180.

Bickmore, T. and Schulman, D. (2007). Practical approaches to comforting users with relational agents. In *CHI '07 Extended Abstracts on Human Factors in Computing Systems* (San Jose, CA, USA, April 28 - May 03, 2007). CHI '07. ACM, New York, NY, 2291-2296.

Chen, W. (2006). Supporting teachers' intervention in collaborative knowledge building. *J. Netw. Comput. Appl.* 29, 2 (Apr. 2006), 200-215.

Cramer, H. S., Evers, V., van Someren, M. W., and Wielinga, B. J. (2009). Awareness, training and trust in interaction with adaptive spam filters. In *Proceedings of the 27th international Conference on Human Factors in Computing Systems*(Boston, MA, USA, April 04 - 09, 2009). CHI '09. ACM, New York, NY, 909-912.

Cramer, H., Evers, V., Ramlal, S., Someren, M., Rutledge, L., Stash, N., Aroyo, L., and Wielinga, B. (2008). The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-Adapted Interaction*18, 5 (Nov. 2008), 455-496.

Dillenbourg, P. (1999). Introduction; What do you mean by Collaborative Learning? In P. Dillenbourg (Ed.) Collaborative Learning. Cognitive and Computational Approaches, (pp. 1-19). Oxford, UK: Elsevier Science.

Fitzpatrick, G., Kaplan, S., Mansfield, T., David, A., and Segall, B. (2002). Supporting Public Availability and Accessibility with Elvin: Experiences and Reflections. *Comput. Supported Coop. Work* 11, 3 (Nov. 2002), 447-474.

Fitzpatrick, G., Mansfield, T., Kaplan, S., Arnold, D., Phelps, T., and Segall, B. (1999). Augmenting the workaday world with Elvin. In *Proceedings of the Sixth Conference on European Conference on Computer Supported Cooperative Work* (Copenhagen, Denmark, September 12 - 16, 1999).

Fuchs, L., Pankoke-Babatz, U., and Prinz, W. (1995). Supporting cooperative awareness with local event mechanisms: the groupdesk system. In *Proceedings of the Fourth Conference on European Conference on Computer-Supported Cooperative Work* (Stockholm, Sweden, September 10 - 14, 1995).

Heiner, J. M., Hudson, S. E., and Tanaka, K. (1999). The information percolator: ambient information display in a decorative object. In *Proceedings of the 12th Annual ACM Symposium on User interface Software and Technology* (Asheville, North Carolina, United States, November 07 - 10, 1999). UIST '99. ACM, New York, NY, 141-148.

Ishii, H., Kobayashi, M., and Arita, K. (1994). Iterative design of seamless collaboration media. *Commun. ACM* 37, 8 (Aug. 1994), 83-97.

Jacob, R. J., Girouard, A., Hirshfield, L. M., Horn, M. S., Shaer, O., Solovey, E. T., and Zigelbaum, J. (2008). Reality-based interaction: a framework for post-WIMP interfaces. In *Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy, April 05 - 10, 2008). CHI '08. ACM, New York, NY, 201-210.

Lee, S. and Lee, W. (2009). Exploring effectiveness of physical metaphor in interaction design. In *Proceedings of the 27th international Conference Extended Abstracts on Human Factors in Computing Systems* (Boston, MA, USA, April 04 - 09, 2009). CHI '09. ACM, New York, NY, 4363-4368.

Manohar, N. R. and Prakash, A. (1994). Replay by Re-execution: A Paradigm for Asynchronous Collaboration via Record and Replay of Interactive Multimedia Sessions. *SIGOIS Bull.* 15, 2 (Dec. 1994), 32-34.

Pedersen, E. R. and Sokoler, T. (1997). AROMA: abstract representation of presence supporting mutual awareness. In*Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, United States, March 22 - 27, 1997). S. Pemberton, Ed. CHI '97. ACM, New York, NY, 51-58.

Shen, H. and Sun, C. (2002). Flexible notification for collaborative systems. In *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work* (New Orleans, Louisiana, USA, November 16 - 20, 2002). CSCW '02. ACM, New York, NY, 77-86.

Soller, A., Martínez, A., Jermann, P., and Muehlenbrock, M. (2005). From Mirroring to Guiding: A Review of State of the Art Technology for Supporting Collaborative Learning. *Int. J. Artif. Intell. Ed.* 15, 4 (Dec. 2005), 261-290.

Tang, J. C. and Rua, M. (1994). Montage: providing teleproximity for distributed groups. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Celebrating interdependence* (Boston, Massachusetts, United States, April 24 - 28, 1994). B. Adelson, S. Dumais, and J. Olson, Eds. CHI '94. ACM, New York, NY, 37-43.

Group Awareness Tools for Controversial CSCL Discussions: Dissociating Rating Effects and Visualized Feedback Effects

Jürgen Buder, Knowledge Media Research Center, j.buder@iwm-kmrc.de
Daniel Bodemer, University of Tübingen, Applied Cognitive Psychology and Media Psychology, Konrad-Adenauer-Str. 40, 72072 Tübingen, Germany, d.bodemer@iwm-kmrc.de

Abstract: An experimental study investigated how a group awareness tool impacts the social influence of a minority faction in controversial online group discussions. The awareness tool involved mutual ratings of contributions on dimensions that make minority opinions more salient. In order to dissociate between the potentially facilitative functions of rating interfaces and visualized feedback, a control condition was compared to a rate-only condition and a tool condition using visualized feedback about ratings. Results indicated that rating without feedback did not strengthen minority viewpoints, but rather widened the differences between minority and majority factions. The group awareness tool that used ratings and visualized feedback yielded moderate effects on minority influence, strong effects on the perceived group preference, and a more pronounced task focus. The results are discussed with regard to the relation of group awareness to different types of social influence and different types of self-awareness.

Introduction

Collaborative learning is first and foremost an activity among peers. Through communication, peers try to achieve shared understanding and to collaboratively construct or re-construct knowledge. Consequently, the role of a teacher fundamentally changes in contexts of collaborative learning. In face-to-face (FTF) collaborative learning, teachers often provide subtle cues to peer communication by providing guiding questions, by eliciting participation, or by applying meta-cognitive strategies of planning and monitoring the interaction. In contexts of purely computer-supported collaborative learning, however, it is not uncommon that teachers do not participate at all in the interaction among learners (e.g. in informal learning scenarios). In these cases, technologies must be designed in order to guide collaboration. There are several ways of how this can be accomplished, e.g. through the use of scaffolds (Scardamalia, 2002) or collaboration scripts (Kollar, Fischer, & Hesse, 2006).

In the last few years the repertoire of CSCL technologies that provide guidance to peer activities has been extended by so-called group awareness tools. The notion of group awareness and the development of tools to foster group awareness have originated in the field of computer-supported cooperative work (Gutwin & Greenberg, 1995). Group awareness tools are technologies that register information about a group, its members and its products, aggregate this information and feed it back to the group members. Originally, group awareness tools were designed to address shortcomings of spatially or temporally distributed group activities. For instance, they sought to re-create the richness of FTF interaction by providing information about the presence of group members in a shared workspace, or by indicating which group member is currently working on which document. Mimicking FTF, however, has become less important once group awareness tools began to be explored in CSCL contexts. One reason for this shift in the conceptualization of CSCL group awareness tools was the general consensus that technologies should provide an added value over FTF scenarios in order to make CSCL justifiable (Buder, 2007). That is, group awareness tools for CSCL should do more than just passively register and feed back "what's going on"; ideally, they should provide guidance to a group and its members. These motivations led to the general idea to develop tools that provide information about a group that would be difficult or even impossible to yield in face-to-face contexts.

For instance, some group awareness approaches in CSCL are based on informing learners about their levels of participation (Janssen, Erkens, & Kirschner, 2011), or require learners to rate the behavior of their collaborators (Phielix, Prins, Kirschner, Erkens, & Jaspers, 2011). However, the most frequently used method involves providing information about the knowledge of collaborators (knowledge awareness tools; Engelmann, Dehler, Bodemer, & Buder, 2009). Knowledge awareness can be conceptualized in very different ways: e.g. by requiring learners to externalize their knowledge by creating concept maps prior to collaboration (Engelmann & Hesse, 2010); by requiring learners to explicitly rate their level of understanding with regard to pieces of learning material (Dehler, Bodemer, Buder, & Hesse, 2011); by making the results of a prior knowledge test available to collaborators at the beginning of interaction (Sangin, Molinari, Nüssli, & Dillenbourg, 2011); or by constraining ongoing collaborative interaction in ways that make differences among learner knowledge visible (Bodemer, 2011). In all these cases, an element that can only be indirectly inferred in FTF interaction (viz. knowledge) is made explicit and salient, thereby guiding collaborative processes. A common finding in the CSCL studies on group awareness is that tools unfold their power by feeding back information about differences among learners. If levels of understanding are different, issues can be resolved by one learner providing explanations to a collaborator; if collaborators differ in how they understand elements of a learning material, they can focus on these conflicting issues and negotiate on a shared understanding.

However, in some learning scenarios resolving conflict might become more complex. This is the case for differences in viewpoints or opinions. An opinion does not relate to isolated arguments, but rather emerges from evaluating and forming a social judgment on a whole set of arguments. This lends a social psychological dimension to collaboration. Conflicts cannot be resolved through explanations or through negotiation of single arguments, but rather involve persuasive communication on how to evaluate a whole set of arguments. Resolving differences in opinion is tricky, as it is likely that individuals process information in a somewhat biased way. For instance, individuals exhibit confirmation bias, a tendency to disregard dissenting information (Jonas, Schulz-Hardt, Frey, & Thelen, 2001). Similarly, in collaborative contexts there is a general tendency for groups not to take the full variability among members into account (Hinsz, Tindale, & Vollrath, 1997). For instance, groups favor shared over unshared information even in contexts where the consideration of unshared information would lead to better group performance (Stasser, 1992). Another facet of variability reduction in groups appears in conflicts between factions of different size, i.e. majorities and minorities. Asch (1951) was the first to show that minority members tend to conform to incorrect viewpoints when they are advocated by a majority.

In order to address these issues, a group awareness tool for resolving different opinions in a group must touch on social psychological dimensions. This idea was at the core of a study about so-called augmented group awareness tools (Buder & Bodemer, 2008). They developed and tested a group awareness tool that tried to resolve learners' different opinions on two conflicting physics hypotheses. Their study was modeled after the informed minority paradigm (Stewart & Stasser, 1998), and it involved 4-person groups where three learners were in favor of a scientifically incorrect hypothesis, and one learner (the informed minority) was in favor of a scientifically correct hypothesis. Learners were required to come to a consensual decision on one of the hypotheses after a 30-minute online forum discussion. The results of this study have shown that minorities tended to conform to the scientifically incorrect majority opinion. However, half of the groups in this study were supported by a group awareness tool that made minority contributions particularly salient. The tool required learners to rate their discussion contributions on two dimensions, viz. agreement with a contribution and perceived novelty of a contribution during discussion. The tool aggregated these ratings, computed average ratings for each contribution, and fed back these results in a visualization where each contribution was represented as a dot in a two-dimensional coordinate system. It was shown that this augmented group awareness tool strengthened minority influence, as groups arrived more often at the scientifically correct minority viewpoint.

While this tool was effective in strengthening minority influence, the underlying mechanisms of the effects deserve further investigation. The augmented group awareness tool rests on two components, both of which might have contributed to its effectiveness. The first component is related to the rating activities. It can be argued that rating a contribution requires reflective thought that would not have occurred to the same degree without rating. In this case, the requirement to rate might have served as a valuable meta-cognitive prompt (Kramarski & Mevarech, 2003) inducing group awareness. The second component that might have contributed to the effectiveness of the tool is the visualized feedback. Only through the visualization, learners could see how the group as a whole thought about the discussion contributions, and only through the visualization minority contributions were becoming salient (lower average agreement and higher average novelty ratings than majority contributions). In order to dissociate between these two components, the study presented here attempted to extend the findings of Buder and Bodemer (2008) by including an experimental condition where learners were asked to rate contributions, but these ratings were neither aggregated, nor fed back to the group members.

It was hypothesized that the components of a rating interface and of visual feedback about ratings affect performance additively. That is, individuals in rate-only groups (with rating, without visualization) should outperform individuals in unsupported groups (no rating, no visualization), and individuals in groups using the complete awareness tool (with rating, with visualization) should outperform individuals in rate-only groups. The ordering of conditions was expected with regard to the following dependent variables:

Hypothesis 1a. Post-discussion preferences leaning towards the minority viewpoint
Hypothesis 1b. Strong minority influence for majority participants (pre- vs. post-discussion)
Hypothesis 1c. Weak majority influence for minority participants (pre- vs. post-discussion)
Hypothesis 1d. Perceived group preferences learning towards the minority viewpoint
Hypothesis 2a. Higher performance in a knowledge test
Hypothesis 2b. Higher knowledge test performance on preference-inconsistent items
Hypothesis 3. Higher rates of discussion focusing on the exchange of arguments
Hypothesis 4. Higher salience of minority contributions through lower agreement ratings, but higher novelty ratings.

To test these predictions, a laboratory experiment was conducted.

Method

The basic setup of this study was adapted from Buder and Bodemer (2008), but the learning domain and the group size were altered. Additionally, this study contained a knowledge test.

Design and Participants

For testing the hypotheses, a one-factorial design with three conditions was employed. In the control condition, groups were using the discussion environment without ratings or visualizations. In the rate-only condition, group members rated discussion posts, but these ratings were not aggregated or fed back to the entire group. In the tool condition, the fully functional group awareness tool was available.

87 student participants (60 female, 27 male; *M* = 24.84 years) were randomly assigned to experimental conditions, and within three-person groups they were randomly assigned to the minority position or one of the two majority positions. Eventually, 29 groups were taking part in the experiment (9 in the control condition; 10 in the rate-only condition; 10 in the tool condition). Subjects were paid for participation, or received course credit. Students of biology or geology were excluded from participation. Prior knowledge of the learning domain was uniformly low across participants (*M* = 2.67, *SD* = 1.24, on a scale ranging from 1 through 7).

Materials

Instructional Material

Instructional material consisted of an introductory text and sets of arguments covering the event that caused the extinction of dinosaurs about 65 million years ago. The introductory text described some basics about the event, and introduced two competing hypotheses about the cause of extinction (meteorite impact vs. long-term volcanism). A pool of ten arguments was created that provided evidence for these hypotheses. Arguments were printed on separate index cards and had an average length of about 160 words. Six of these arguments were in favor of the volcanism hypothesis, whereas only four arguments were in favor of the meteorite hypothesis, thus making the volcanism hypothesis superior. At the time of testing, the actual merits of both hypotheses were still hotly debated among geologists and paleontologists. However, popular scientific accounts had much stronger coverage of the meteorite hypothesis, thereby ensuring a conservative testing of social influence effects. Prior tests had shown that nine out of ten subjects who received the whole set of arguments were favoring the volcanism hypothesis afterwards. For the experiment, two sets of arguments were created. One set was identical for all majority members and consisted of all four arguments favoring the meteorite hypothesis plus two arguments favoring the volcanism hypothesis. The set for the minority members consisted of all six arguments favoring the volcanism hypothesis plus two arguments favoring the meteorite hypothesis.

Online Discussion Environment

The online discussion environment was developed at the Knowledge Media Research Center in Tübingen. The environment consists of a temporally ordered list of separate posts. Author names are anonymized. Depending on experimental conditions, a rating interface and an awareness tool were available. The rating interface was implemented as two sliders attached to each discussion post except one's own. One slider was designated to express novelty ratings; the other was designated to express agreement ratings. The ratings on the sliders were expressed as numbers between 0 and 100. The awareness tool contained a visualization of the discussion posts represented as dots on a two-dimensional graph, where the x-axis represented the average agreement rating, and the y-axis represented the average novelty rating that a given contribution received. The visualization was personalized in that learners could distinguish their own contributions from other group members' contributions, and by indicating contributions a learner hadn't rated yet (see Figure 1). By clicking on a particular dot in the visualization the discussion window automatically scrolled to the selected discussion post.

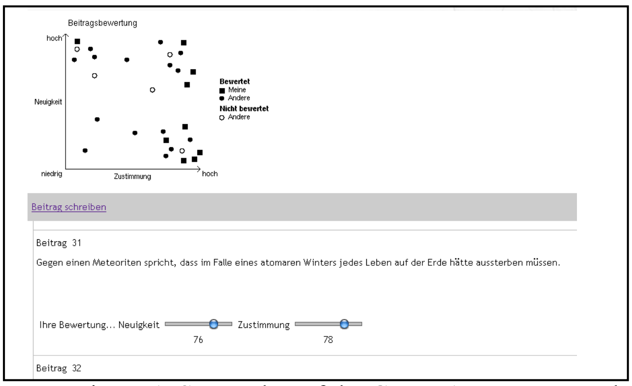


Figure 1. Screenshot of the Group Awareness Tool.

Measurements

In order to test the general predictions concerning the effectiveness of ratings and visualizations, four classes of dependent measures were analyzed. The first two classes refer to outcome variables (preference data, knowledge test data); the third and fourth classes refer to process variables (participation data, rating data).

Preferences were captured by measuring learner ratings on corresponding sliders ranging from -100 ("complete agreement with meteorite hypothesis") to +100 ("complete agreement with volcanism hypothesis"). Sliders were used to capture individual pre-preference, individual post-preference and perceived group preference. Minority influence and majority influence were measured by computing the difference from pre- to post-preference in the direction of the scale midpoint.

Knowledge test performance was measured as number of correct responses in a 10-item multiple-choice test (with one target and three distractors, respectively).

Participation data were captured both through objective data (number of posts) and through content analysis. For the content analysis, all posts were coded by two independent raters. Among other things, it was coded for each post whether it contained an explanation of an argument (Cohen's *kappa* = .88).

Finally, *rating* data on average agreement and average novelty were captured for each post in the rate-only condition and the tool condition.

Procedure

The experiment consisted of three phases: an individual learning phase, a group discussion phase, and an individual knowledge test phase. During the entire experiment subjects of a group were seated in separate rooms. In the first phase learners individually worked through the learning material on the dinosaur extinction event (20 minutes). After that, individual pre-discussion preferences were measured. After the learning phase, group members were given the opportunity to test the online discussion environment by writing some contributions. In the rate-only and tool conditions, participants were asked to rate test contributions by other learners. In the tool condition, the functionality of the awareness tool was explained during this stage.

In the second phase index cards containing the arguments were removed, and groups were instructed to discuss the conflicting hypotheses using the online environment. All learners were made aware that other group members might have received different pieces of evidence. Groups were asked to come to an agreement about the conflicting hypotheses within the allotted discussion time (30 minutes). According to the experimental design of the study groups in the control condition were only provided with the online discussion environment. Groups in the rate-only condition were additionally asked to rate contributions of their collaborators on agreement and novelty (e.g. low ratings for arguments that are mentioned repeatedly) by using two sliders ranging from 0 to 100. Groups in the tool condition also used the rating interface, but were additionally provided with visualized feedback. After the discussion phase individual learners were asked to indicate their post-discussion preference and their perceived overall group preference.

In the third phase, participants individually worked through the multiple-choice knowledge test. Subjects were briefed about the study at the end of the experiment.

Results

Means and standard deviations for *preference data* are listed in Table 1. The experimental manipulation was checked using a two-factorial analysis of variance (ANOVA) with condition (control, rate-only, tool) and status (majority, minority) as factors and *pre-discussion preference* as dependent variable. As expected, it yielded no main effect for condition ($F = 0.24$, $p = .78$, $\eta^2 = .00$); a main effect for status: $F(1, 81) = 399.90$, $p < .01$, $\eta^2 = .73$; and no interaction effect: $F(2, 81) = 1.55$, $p = .22$, $\eta^2 = .01$. This indicates that the manipulation worked properly.

Table 1: Means and standard deviations for condition and status with regard to preference data.

Condition	Status	Pre-preference	Post-preference	Influence	Group preference
Control	Majority	-52.3 (24.2)	-38.6 (49.7)	13.8 (53.3)	-39.9 (52.6)
Rate-only	Majority	-61.1 (25.0)	-64.6 (39.8)	-3.5 (31.0)	-46.4 (39.0)
Tool	Majority	-61.7 (29.7)	-40.5 (45.2)	21.3 (47.2)	-11.7 (45.6)
Control	Minority	46.8 (16.8)	30.9 (28.8)	15.9 (27.7)	-44.8 (49.5)
Rate-only	Minority	62.2 (19.3)	69.5 (22.8)	-7.3 (27.8)	-28.3 (25.5)
Tool	Minority	53.6 (24.3)	53.8 (26.3)	-0.2 (17.6)	-3.5 (52.7)

Post-discussion preferences indicate in how much learners tended towards the correct minority viewpoint (Hypothesis 1a). These analyses were conducted using the same 3 (condition) x 2 (status) analysis of variance as for pre-discussion preferences. The ANOVA did not yield the expected main effect for condition: $F(2,81) = 0.47$, $p = .63$, $\eta^2 = .00$); the strong main effect for status remained: $F(1, 81) = 126.05$, $p < .01$, $\eta^2 =$

.54. However, there was a significant interaction effect: $F(2, 81) = 4.50$; $p = .01$, $\eta^2 = .04$. Additional analyses revealed that this interaction was due to the large majority-minority spread in the rate-only condition. It appears that rating without subsequent visualized feedback actually enforces initial preferences of both minorities and majorities.

Minority influence (Hypothesis 1b) was measured using a one-factorial ANOVA for majority participants with regard to difference between post- and pre-preferences. The data reveal that minority influence was strongest in the tool condition, but the effect failed to reach significance: $F(2, 55) = 1.62$; $p = .21$, $\eta^2 = .05$.

A similar analysis was conducted for (undesirable) *majority influence* (Hypothesis 1c), based on minority participants. Majority influence was strongest in the control condition, but again, the effects were not strong enough to yield significance: $F(2, 26) = 2.17$; $p = .13$, $\eta^2 = .14$.

Participants were also required to express *perceived group preference* (Hypothesis 1d). The corresponding analyses were conducted with a 3 (condition) x 2 (status) analysis of variance. Here, a significant main effect for condition could be obtained: $F(2, 81) = 4.41$, $p = .02$, $\eta^2 = .07$. A post-hoc test revealed that participants in the tool condition perceived the group decision to be closer to the minority viewpoint than participants in the other two conditions. Neither a status effect ($F(1,81) = 0.46$, $p = .50$, $\eta^2 = .00$) nor a significant interaction ($F(2,81) = 0.39$, $p = .68$, $\eta^2 = .00$) were obtained.

Table 2: Means and standard deviations for condition and status with regard to knowledge test performance, participation, and rating data.

Measurement	Status	Control	Rate-Only	Tool
Knowledge test overall	Majority	.73 (.21)	.63 (.14)	.68 (.11)
	Minority	.73 (.10)	.77 (.15)	.79 (.11)
Knowledge on preference-inconsistent items	Majority	.66 (.24)	.54 (.19)	.58 (.18)
	Minority	.52 (.29)	.70 (.19)	.63 (.11)
Number of contributions	Majority	17.9 (9.8)	12.8 (5.6)	11.2 (5.8)
	Minority	22.0 (5.9)	13.9 (5.3)	11.1 (3.2)
Relative number of explanations	Majority	.29 (.14)	.44 (.19)	.56 (.21)
	Minority	.43 (.19)	.57 (.18)	.74 (.22)
Received agreement ratings	Majority		61.8 (14.4)	68.1 (14.7)
	Minority		43.1 (11.3)	50.7 (14.0)
Received novelty ratings	Majority		56.3 (15.8)	50.3 (16.4)
	Minority		67.9 (11.8)	65.0 (12.1)

A secondary outcome variable was the *knowledge test* (see Table 2). For *overall performance* (Hypothesis 2a), no significant main effect for condition could be found: $F(2,81) = 0.38$, $p = .68$, $\eta^2 = .00$. There was a very small, but significant main effect for status: $F(1, 81) = 5.74$, $p = .02$, $\eta^2 = .00$, with minority members showing better test performance, but this was probably due to the fact they initially received more information about the arguments than majority members. The condition x status interaction was $F(2, 81) = 1.42$, $p = .25$, $\eta^2 = .00$.

As to subsets of the knowledge test, additional analyses were conducted with the performance on *preference-inconsistent items* (Hypothesis 2b). While no main effects for condition ($F(2, 81) = 0.16$, $p = .85$, $\eta^2 = .00$) or status ($F(1, 81) = 0.31$, $p = .58$, $\eta^2 = .00$) were observed, the data revealed a significant interaction: $F(2, 81) = 3.43$, $p = .04$, $\eta^2 = .01$. In the control condition, majority members outperformed minority members, whereas this pattern was reversed in the rate-only condition. The tool condition yielded similar performance levels for majorities and minorities. As the control condition had the strongest majority influence, and the rate-only condition did not show any minority influence (see Table 1), the results hint at the possibility that performance on preference-inconsistent items might be related to patterns of social influence. Additional analyses revealed that over all three conditions, majority influence was negatively correlated with performance on preference-inconsistent items ($r = -.10$, $p = .60$), whereas minority influence was positively correlated with performance on preference-inconsistent items ($r = .24$, $p = .07$). While these correlations were non-significant, a tendency could be observed that being influenced by a majority was associated with less learning of majority-related concepts, whereas being influenced by a minority led to higher learning of minority-related concepts.

The *participation data* indicated that subjects in the control condition wrote more *contributions* than group members in the other conditions: $F(2, 81) = 12.11$, $p < .01$, $\eta^2 = .05$. Neither a main effect for status – $F(1, 81) = 1.34$, $p = .25$, $\eta^2 = .05$ - nor a condition x status interaction was found ($F(2, 81) = 0.66$, $p = .52$, $\eta^2 = .00$). A more detailed result was revealed through the analysis of the *relative amount of explanations* (Hypothesis 3). Here, a significant main effect for condition was found: $F(2, 81) = 15.07$, $p < .01$, $\eta^2 = .04$. Post-hoc tests revealed that in the tool condition, significantly more explanations occurred than in the rating-only condition, and in both these conditions, more explanations occurred than in the control condition. Moreover, minority members produced a higher rate of explanatory posts than majority members; $F(1,81) = 12.09$, $p < .01$, $\eta^2 = .02$. The interaction was non-significant: $F(2,81) = 0.17$, $p < .84$, $\eta^2 = .00$.

Finally, *rating data* for the rating-only and the tool condition were analyzed (Hypothesis 4). About 90% of contributions were rated by participants in both conditions. For the *agreement ratings*, a marginally significant main effect for condition was found: $F(1, 56) = 3.28$, $p = .08$, $\eta^2 = .00$, indicating that participants in the tool condition expressed slightly higher agreement with the discussion posts than participants in the rate-only condition. As expected, majority contributions generally received higher agreement ratings than minority contributions; $F(1, 56) = 22.29$, $p < .01$, $\eta^2 = .02$. No interaction effect was obtained: $F(1, 56) = 0.03$, $p = .87$, $\eta^2 = .00$. As for *novelty ratings*, conditions did not differ significantly from each other: $F(1, 56) = 1.18$, $p = .28$, $\eta^2 = .00$; however, minority contributions received higher novelty ratings than majority contributions: $F(1, 56) = 10.44$, $p < .01$, $\eta^2 = .01$. Again, the interaction was non-significant: $F(1, 56) = 0.15$, $p = .70$, $\eta^2 = .00$.

Discussion

The present study investigated the influence of rating interfaces and visualized feedback of an awareness tool on individual preferences, perceived group preferences, and individual knowledge test performance. It was expected that ratings serve as meta-cognitive prompts that improve outcome variables. Moreover, a visualized feedback about average ratings through an awareness tool was expected to enhance these positive effects. The predictions were tested in an experimental design involving three conditions (control groups, rate-only groups, tool groups).

The preference data yielded rather mixed results. For post-discussion preferences it was not found that participants in the tool condition were leaning stronger towards the minority viewpoint. Patterns of social influence revealed that unsupported groups exhibited strong majority influence and a moderate minority influence. In the tool condition, majority influence was non-existent whereas minority influence was highest among conditions. While this result is in line with predictions, the effects were too weak to reach significance. In contrast, rate-only groups showed neither minority influence nor majority influence. Rather, requesting repeated ratings without any feedback appeared to strengthen initial individual preferences. For this reason, the explanatory mechanism of meta-cognitive stimulation through ratings must be ruled out for this scenario. The data on perceived group preferences are more encouraging. Participants that were supported by an awareness tool estimated the average preference of their groups stronger in favor of the volcanism hypothesis than participants in the other two conditions. This can be interpreted as a higher awareness for the minority opinion. However, it should be noted that in this scenario the meteorite hypothesis was generally deemed much stronger. Only in one out of nine control groups, one out of ten rate-only groups, and three out of ten tool groups the averages of perceived group preferences tended towards the volcanism hypothesis. This might be due to the fact that in popular scientific accounts of the extinction event the meteorite hypothesis has received much higher coverage.

The knowledge test data did not yield overall differences among the three conditions, thus the hypothesis that ratings and rating visualizations increase performance cannot be confirmed. The detailed results for test performance on preference-inconsistent items revealed some interesting effects: in conjunction with the data on social influence pattern it was found that majority influence was associated with lower performance on majority test items. In other words, minorities might shift towards the majority opinion, but learn relatively little about that opinion. Such a pattern could be interpreted as normative social influence, an unthinking adoption of the majority viewpoint. In contrast, minority influence was associated with higher performance on minority test items. This indicates that majority members who shift towards the minority opinion learn more about that opinion. This higher performance might be due to informational social influence, an effect that causes majorities to scrutinize minority viewpoints more carefully (Wood, Lundgren, Ouellette, Busceme, & Blackstone, 1994).

The participation data can be interpreted in a way that ratings and visualized feedback about ratings both improve the learning process. The overall reduction of posts in the rate-only and tool conditions was probably due to the fact that rating and/or using the visualization takes time, thereby leading to lower productivity. However, this was offset by a higher task focus. In the tool condition, the rate of explanatory posts was significantly higher than in the rate-only condition, and the latter condition in turn yielded a higher rate of explanatory posts than the control condition. Additional analyses revealed that control groups generated a much higher amount of posts pertaining to task coordination and off-topic talk. A possible interpretation for this pattern is that rating activities subtly structure individual learning processes, thereby reducing the need for explicit coordination among collaborators.

Rating data might help to illuminate the social influence and deliberation processes of rate-only groups vs. tool groups. While the rating behavior of rate-only groups did not differ much from tool groups, the data indicate that minority contributions were rated lower on agreement, but much higher on novelty than majority contributions. As a consequence, minority contributions became visually salient in the tool condition which is exactly what the augmented group awareness tool tried to capture. Moreover, it was found that tool groups expressed slightly higher agreement with the discussion posts which might have contributed to an atmosphere where majority members scrutinized minority contributions more carefully and thus experienced stronger minority influence. It might be the case that additional analyses on subsets of posts will yield even stronger effects. In the current study, participants were asked to rate each discussion post, and it can be argued that agreement with a question cannot be interpreted in the same way as agreement with an answer. This also begs the question of whether the effectiveness of awareness tools and the clarity of experimental results can be improved by instructing collaborators to rate only those contributions that they regard as essential for the learning process.

Taken together, the results indicate that a rating interface per se is not sufficient to improve collaborative learning. The fact that social influence was virtually non-existent in the rate-only condition suggests that rating without feedback does not foster reflection on other group members and their products. On the contrary, it might be the case that repeated ratings direct attention to the self, thereby strengthening initial preferences. In other words, rating without feedback might not evoke group awareness, but *private self-awareness*, a tendency to adhere to *personal* standards of behavior (Froming, Walker, & Lopyan, 1982).

The results also indicate that visualized feedback about ratings can lead to social influence. In a scenario where the minority opinion is associated with better outcomes, the tool condition led to a moderate minority influence, and practically no majority influence. Moreover, participants in the tool condition perceived the group preference to be more shifted towards the minority opinion, exhibited more explanatory behavior, and expressed slightly higher agreement with the contributions of others. In this regard, it can be said that group awareness was achieved through the tool. However, since the visualization also provided feedback about oneself, it can be argued that the awareness tool also increased *public self-awareness*, the tendency to adhere to *social* standards of behavior (Froming et al., 1982).

Conclusions

The objectives of the present study were twofold. First, further evidence for the effectiveness of group awareness tools to foster computer-supported collaborative learning should be obtained. Social influence patterns were weaker than in the study by Buder and Bodemer (2008). This might be due to the selected learning domain, as argumentation skills can vary considerably across different domains (Mason & Scirica, 2006). Moreover, inducing preference change is difficult in a limited time frame. Another explanatory mechanism for the differences between the original study and the replication study could be subtle differences in the framing of the task that led to a differentiation-focused debate mode rather than an integration-focused controversy mode (Johnson & Johnson, 1979). Future studies could address how the awareness tool used here would work in a setting where the task focus is on open discussion rather than on joint decision making. Moreover, it would be interesting to see how the tool influences collaborative learning in a scenario where majority and minority factions are either non-existent or randomly distributed.

There have been many empirical studies showing the effectiveness of group awareness tools in CSCL scenarios. However, relatively little work has been done to uncover the mechanisms of group awareness (Buder, 2011). The second objective of the present study was to contribute to our understanding of *how* these tools actually work. This was attempted through the inclusion of an experimental rate-only condition that was not inspired by principles in the learning sciences, but served mainly to dissociate rating effects from visualized feedback effects. The results do not only indicate that isolated ratings are detrimental to collaborative learning, but also help to establish links between normative majority influence, informational minority influence, private self-awareness, public self-awareness, and group awareness.

References

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership, and men* (pp. 177-190). Pittsburgh: Carnegie Press.

Bodemer, D. (2011). Tacit guidance for collaborative multimedia learning. *Computers in Human Behavior, 27*(3), 1079-1086.

Buder, J. (2007). Net-based knowledge-communication in groups: Searching for added value. *Zeitschrift für Psychologie/Journal of Psychology 215*(4), 209-217.

Buder, J. (2011). Group awareness tools for learning: Current and future directions. *Computers in Human Behavior, 27*(3), 1114-1117.

Buder, J., & Bodemer, D. (2008). Supporting controversial CSCL discussions with augmented group awareness tools. *International Journal of Computer-Supported Collaborative Learning, 3*(2), 123-139.

Dehler, J., Bodemer, D., Buder, J., & Hesse, F. W. (2011). Guiding knowledge communication in CSCL via group knowledge awareness. *Computers in Human Behavior, 27*(3), 1068-1078.

Engelmann, T., Dehler, J., Bodemer, D., & Buder, J. (2009). Knowledge awareness in CSCL: A psychological perspective. *Computers in Human Behavior 25*(4), 949-960.

Engelmann, T., & Hesse, F.W. (2010). How digital concept maps about the collaborators' knowledge and information influence computer-supported collaborative problem solving. *International Journal of Computer-Supported Collaborative Learning, 5*(3), 299-320.

Froming, W. J., Walker, G. R., & Lopyan, K. J. (1982). Public and private self-awareness: When personal attitudes conflict with societal expectations. *Journal of Experimental Social Psychology, 18*(5), 476-487.

Gutwin, C., & Greenberg, S. (1995). Support for group awareness in real-time desktop conferences. In: *Proceedings of the Second New Zealand Computer Science Research Students' Conference.* (University of Waikato, Hamilton, New Zealand), April 18-21, 1995.

Hinsz, V. B., Tindale, R. S., & Vollrath, D. A. (1997). The emerging conceptualizion of groups as information processors. *Psychological Bulletin, 121*, 43-64.

Janssen, J., Erkens, G., & Kirschner, P. A. (2011). Group awareness tools: It's what you do with it that matters. *Computers in Human Behavior, 27*(3), 1046-1058.

Johnson, D. W., & Johnson, R. T. (1979). Conflict in the classroom: Controversy and learning. *Review of Educational Research*, *49*(1), 51-69.

Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology*, *80*(4), 557-571.

Kollar, I., Fischer, F., & Hesse, F. W. (2006). Collaboration scripts - a conceptual analysis. *Educational Psychology Review, 18*(2), 159-185.

Kramarski, B., & Mevarech, Z. R. (2003). Enhancing mathematical reasoning in the classroom: Effects of cooperative learning and metacognitive training. *American Educational Research Journal, 40*(1), 281-310.

Mason, L., & Scirica, F. (2006). Prediction of students' argumentation skills about controversial topics by epistemological understanding. *Learning & Instruction, 16*(5), 492-509.

Phielix, C., Prins, F. J., Kirschner, P. A., Erkens, G., & Jaspers, J. (2011). Group awareness of social and cognitive performance in a CSCL environment: Effects of a peer feedback and reflection tool. *Computers in Human Behavior, 27*(3), 1087-1102.

Sangin, M., Molinari, G., Nüssli, M.-A., & Dillenbourg, P. (2011). Facilitating peer knowledge modeling: Effects of a knowledge awareness tool on collaborative learning outcomes and processes. *Computers in Human Behavior, 27*(3), 1059-1067.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal education in a knowledge society* (pp. 67-98). Chicago: Open Court.

Stasser, G. (1992). Pooling of unshared information during group discussions. In S. Worchel, W. Wood. & J. A. Simpson (Eds.), *Group process and productivity* (pp. 48-67). Newbury Park: Sage

Stewart, D.D. & Stasser, G. (1998). The sampling of critical, unshared information in decision-making groups: The role of an informed minority. *European Journal of Social Psychology, 28*, 95-113.

Wood, W., Lundgren, S., Ouellette, J. A., Busceme, S., & Blackstone, T. (1994). Minority influence: A meta-analytic review of social influence processes. *Psychological Bulletin, 115*(3), 323-345.

Missing Something? Authority in Collaborative Learning

Iris Howley, Elijah Mayfield, Carolyn P. Rosé, Carnegie Mellon University, 5000 Forbes Ave. Pittsburgh, PA
Email: ihowley@cs.cmu.edu, emayfiel@cs.cmu.edu, cprose@cs.cmu.edu

Abstract: Past research in individual learning settings has shown that student dispositions such as self-efficacy are predictive of learning and other beneficial outcomes, but the relationship is less clear in a collaborative learning environment. This paper explores authoritativeness of stance within a conversation as a social factor influencing learning and related to self-efficacy in a computer-supported collaborative learning setting. Our results indicate that this authoritativeness measure predicts learning, where an individual's self-efficacy does not, and that student and partner authoritativeness predicts group self-efficacy. Further research is required to better determine the relationship between conversational authoritativeness, individual dispositions, and learning.

Introduction

Research in individual learning settings connected with popular dispositional constructshas demonstrated that intrinsic motivation, mastery-learning oriented goals, and high self-efficacy are predictive of learning as well as positive traits such as persistence (Sheldon & Kasser, 1995; Harackiewicz et al, 2002; Coutinho & Newman, 2008). However, this leaves open the question of what happens when an additional student is added to the learning activity, and we then have a collaborative learning task. Will self-efficacy still predict learning and behavior in this collaborative setting? Or does the absence of considered social factors in these dispositional constructs dilute their predictive value in the face of social comparison, social identity, positioning, and other factors introduced by a collaborative setting?

At the same time that the introduction of a learning partner introduces complications that may interfere with the predictive value of motivational assessments largely optimized for use within individual learning settings, the advantage is that the social interaction makes self-concepts visible through linguistic strategies speakers employ to position themselves within their groups. Thus, as an important part of our work investigating the applicability of constructs such as self-efficacy within a group learning setting, we develop a behavioral measure in the form of a categorical coding scheme through which we can quantify the relative level of authoritativeness of stance between the collaborative partners. In this paper we apply this coding scheme to conversational data by hand, although in the long term, our hope is to be able to automate its application by means of machine learning technology as has been used in prior work on collaborative learning process analysis (Rosé et al., 2008; Mayfield & Rosé, 2011). In this paper we explore the relationship between this proposed behavioral measure of authoritativeness and dispositional attributes such as self-efficacy. If behavioral measures of authoritativeness (or some other automated coding scheme) can be used to predict student learning or persistence, and if these measures can be automated, then the Computer-Supported Learning (CSCL) research community could leverage some of the benefits of individual disposition research, automatically assessed through chat contributions, in naturalistic collaborative learning environments. For example, if we could automatically measure self-efficacy through chat behavior, we may not need to use self-report questionnaires. Self-report questionnaires, while useful in a research atmosphere, are not as plausible in more natural CSCL environments where there may not be an experimenter, or even a permanent instructor.

In the remainder of this paper, we begin by outlining the prior research on assessment of dispositional attitudes. Next, we introduce our operationalization of "authoritativeness" as a behavioral measure capturing one factor that may exist on the social dimension of collaborative learning. We use our measure of authority in knowledge along with self-efficacy to examine learning in a collaborative context as a reanalysis of a dataset from an earlier computer supported collaborative learning study (Ai et al., 2010). We conclude with a discussion of our current directions and future work.

Prior Work

Much prior work has examined the effects of dispositional attributes on learning in individual contexts. In general, measures of intrinsic motivation have been associated with a variety of positive implications. For example, studies have shown that students pursuing intrinsic goals are not only intrinsically motivated, but also portray behaviors enhancing their well-being (Sheldon & Kasser, 1995). Both the pursuit of intrinsic or extrinsic goals and the possession of autonomous or external motives have independent effects on well-being (Sheldon et al., 2004). Additional research based upon help seeking and achievement goal theory (Eccles & Wigfield, 2002; Dweck, 1986; Nicholls, 1984; Harackiewicz et al., 2002) shows that an intrinsic goal-orientation reduces help seeking avoidance and increases the likelihood of more optimal help seeking strategies (Newman, 1990; Ryan & Pintrich, 1997). Self-efficacy is a strong predictor of learning and motivation in individual environments

(Zimmerman, 1999), which strongly suggests it may have invaluable potential for collaborative environments. In this paper we will focus on Bandura's (1977) theory of self-efficacy in a thermodynamics collaborative task. Academic self-efficacy is a student's perceptions of her academic capabilities, interpreted from previous mastery experience, vicarious experience, verbal and social persuasions, and emotional and physiological states. Self-efficacy beliefs contribute to the choices students make, as well as their persistence and effort expended.

Despite this substantial body of positive results, these constructs would seem logically to have implications within collaborative learning settings as they have been demonstrated to have in individual learning settings, but little work has investigated how these traits influence behaviors within those social contexts. When these dispositions are examined in collaborative settings, the picture is far more complex. These measures sometimes continue to predict learning, but often predict something else at the same time. For example, Darnon et al (2006) shows that differing achievement goal orientations result in different approaches to mediating conflict through epistemic regulation for mastery-oriented students and through relational regulation for performance-oriented students. However, beyond this, little work has been done to unite self-efficacy with collaborative learning. One explanation for the murkiness of the relationship between individual dispositions and collaborative learning is that two students collaborating together is more than just two individuals with a common task. The same discussion that produces the cognitive conflict leading to learning (Piaget, 1985) adds a dynamic social dimension to the learning activity that may introduce other factors that interfere with the causal relationship between individual dispositions and learning.

For group situations, collective efficacy has been proposed as an alternative. According to Bandura (1997), collective efficacy is several individuals' combined perception of the group's capabilities to perform given tasks. As an example of its application, Wang & Lin (2007) examine self-efficacy in a collaborative learning task, where they incorporate collective efficacy to determine how the self-efficacy configuration of three-person groups predicts collective efficacy and usage of high-level cognitive skills during discussion. While Wang & Lin (2007)'s results suggest a relationship between homogenous collective efficacy, high-level cognitive skills, and group performance, the results are less clear for heterogeneous self-efficacy groups.

In this paper, we propose a conversational analysis framework for encapsulating the social positioning dimension of collaborative tasks, with the goal of eventually automating the process of identifying instances of social shift. We introduce an abbreviated framework for identifying authority in dialogue, which we later use to examine social positioning.

The Authoritativeness Framework

It is reasonable to believe that the social interactions that occur during a collaborative task influence learning and the effect that dispositional attributes have on learning. So, in addition to examining traditional dispositions (in this study, self-efficacy) via self-report, we propose a framework for looking at authoritativeness of knowledge presentation. In short, an authoritative presentation of knowledge is one that is presented without seeking external validation for the knowledge.

The Authoritativeness Framework we introduce in this paper is rooted in Martin's Negotiation Framework (Martin, 1992), from the systemic functional linguistics community. This work highlights the moves that are made in a dialogue as they reflect the authoritativeness with which those moves were made, and gives structure to exchanges back and forth between participants. Previous work has studied the complexity of, for instance, the difference between authority to alter the direction of a conversation and authority to contribute new information to a conversation (Martin, 2000). In its use within the systemic functional linguistics community, it has been used as a way to distinguish between classrooms where the reasoning of the teacher is at center stage from those where student reasoning is the focus (Veel, 1999). It has also been used to investigate subtleties about distribution of power in juvenile trials (Martin et al., 2008), in keeping with the emphasis in the systemic functional linguistics community for using analysis of language to support social justice. We are interested in this framework because of its descriptiveness for social interactions, and how it boils down the intricacies of power management within an interaction to a few simple codes, making it easy to track shifts in positioning over time.

While the Negotiation framework as formulated by Martin is highly descriptive for sociolinguists, and has been widely used by Martin himself as well as by other sociolinguistics, it is difficult to replicate reliably from the previously published formulations, as this was not a methodological goal of the original researchers. This makes its immediate use for quantitative analysis difficult without introducing threats to internal validity. To remedy this, we have worked iteratively on a coding manual that incorporates the insights from that framework that are relevant to our task and makes them precise and concrete enough to be reproducible. Our inter-rater agreement for this coding has achieved a Cohen's Kappa of 0.78. A full treatment of the details of our development process is beyond the scope of this paper, but is discussed more in depth in Howley et al (2011). We would like to acknowledge that we developed this Authoritativeness Framework through consultation with experts from a variety of backgrounds (sociolinguists, computational linguistics, computer scientists, interaction analysts, learning scientists, sociocultural and education researchers, etc).

Our formulation of the Authoritativeness framework is comprised of two dimensions with six and three codes, respectively, and incorporates structural and pragmatic knowledge of language based on the Negotiation framework. To simplify our analysis for this paper, we will focus on two moves in particular. The first is K1, or 'primary knower', and the second is K2, or 'secondary knower'. A 'primary knower' move includes a statement of fact, an opinion, or an answer to a factual question, such as 'yes' or 'no'. It only counts as 'primary knower' if it is not presented in such a way as to elicit an evaluation from another participant in the discussion. Conversely, a 'secondary knower' move includes statements where the speaker is not positioned as authoritative on the topic at hand, such as asking a question eliciting information, or presenting information in a context where evaluation is the expected response or formulated in such a way as to elicit feedback. A brief overview of the codes from our Authoritativeness framework is shown in Figure 1.

Code	Meaning	Example
K1	Primary Knower	"This is the end."
K2	Secondary Knower	"Is this the end?" and "This is the end, right?"
A1	Primary Actor	"I'm going to the end."
A2	Secondary Actor	"Go to the end."
ch	Challenge	"I don't have an end marked."
o	Other	"So…"

Figure 1. An Overview of the Codes Included in our Authoritativeness Framework.

There is no strict form-function relationship between these codes and the text being analyzed. The simplest example of this is a line such as 'yeah', which could be authoritative in response to a question or could be non-authoritative response to someone else's evaluation. Additionally, factual statements where the speaker is uncertain of their correctness and is looking for approval from a listener would be coded as a K2 move, even though it is structurally similar to most K1 moves. The roles that speakers take through these codes can shift rapidly within a conversation, and are dynamic, being heavily based on the context of what has happened leading up to an utterance, and how that utterance is responded to by other participants.

	Name	Text	Authority
1	**Student1**	any idea what we open to start?	K2
2	**Student2**	I just opened the reheat cycle.	K1
3	**Student2**	It's blank, and it's already a system.	K1
4	**Student2**	We can substitute values in as we go.	K1
5	**Student1**	hm okay	o
6	**Student1**	do you understand how to maximize the efficiency of a reheat cycle?	K2
7	**Student2**	I think I do, but since we have opposing goals I'll be helping you along to a point. =)	K1
8	**Student2**	Anyway, we have constants that we have to deal with such as the material being SS and the fluid being water	K1
9	**Student1**	woah woah woah	o
10	**Student1**	what's your goal?	K2
11	**Student1**	efficiency?	K2
12	**Student2**	No, no, I'm going for green	K1
13	**Student2**	Yay team.	o
14	**Student2**	Anyway, did you want to do reheat or simple?	K2
15	**Student2**	You seemed to have reservations before.	o
16	**Student1**	I think reheat.	K1
17	**Student1**	I'll explore the simple one quickly.	A1
18	**Student2**	By the way, I take it your goal is power?	K2
19	**Student1**	yup	K1
20	**Student2**	Great...	o

Figure 2. An Example Analysis Using Martin & Rose's (2003) Negotiation System, Labeled as Authority.

For example, looking at Figure 2, we will see that the roles of primary and secondary speakers is highly volatile and does not appear to carry any particular lasting significance with respect to status distribution within the conversation. Rather than signify any persistent difference in status distribution between the two students in Figure 2, this frequent shifting in speech roles serves to underscore the equal footing between the two students despite the fact that Student2 is clearly more knowledgeable. Furthermore, speech roles are meaningful even where transitory in that they signify which speakers are treated as the source and recipient respectively of the information or goods and services being exchanged. Thus, it allows us to ask not only which speakers are cast as authoritative within an interaction, but authoritative with respect to what.

We can examine the working dynamics of group members and compare what group behaviors are visible in teams with different compositions of authoritativeness and learning gain. Groups where both students are highly authoritative share characteristics. Their exchanges largely consist of statements of fact or proposals for answers to questions from the tutor. Both students appear to understand the concepts that they are discussing, and they often come to similar conclusions. Thus, there is little debate before coming to a consensus.

Figure 3 is an example of a group in which both students are highly authoritative. The only secondary knower (K2) move, line 2, is a proposal for an answer to the tutor, albeit one that is phrased to ask for evaluation from the other student. The rest of the moves are assertions. While there is a disagreement between lines 1 and 2, the justification of each student's stance are given as factual statements (at lines 3 and 5), and the consensus is phrased by each student as if they are authorizing the final decision (lines 7 and 8).

	Name	Text	Authority
1	Student5	alright so around 11,000 kPa	K1
2	Student6	So that would be, like, 14,000 kPa?	K2
3	Student6	Because right after that your graph starts going down faster	K1
4	Student5	Yea	K1
5	Student5	But the peak is a little higher towards 11,000	K1
6	Student5	But whatevs	-
7	Student5	I'll take 14,000	K1
8	Student6	Cool. 14,000 kPa it is.	K1

Figure 3. A Sample of Discussion from Two Students with High Authoritativeness Ratios.

In other common situations, however, the authoritativeness of the two students is very distinct, with one student taking on an authoritative tone and the other being much more submissive. This usually comes in the form of one student asking 'permission' when suggesting new ideas, and the other student making a habit of affirming those questions and giving new information as primary knower (K1) moves.

	Name	Text	Authority
1	Student9	you want lowest waste heat…	K1
2	Student9	So 20,000 kpa?	K2
3	Student9	and mine is that peak at the net power output	K1
4	Student8	need to watch out for the quality	K1
5	Student8	Oh	-
6	Student9	So we want the same pressure then basically?	K2
7	Student8	cant have pmax more than 6,574ish	K1
8	Student8	but that could change based on the temperature coices	K1
9	Student9	So…like…	-
10	Student9	6800? Or something?	K2
11	Student8	That's probably about right	K1

Figure 4. A Chat Example between Two Students with Unequal Amounts of Authoritativeness.

Above, in Figure 4, is an exchange with an unequal ratio of authority, a common pattern. Despite the fact that both students are proposing new ideas, one student repeatedly phrases them without authority (lines 2, 6, and 10). In this case, since the more authoritative student disagrees with these proposals (lines 4 and 7), the tone of the student becomes less authoritative over time. Meanwhile, the more authoritative student has taken on a tutor-like role, confirming or rejecting suggestions from the less authoritative student as seen on line 11.

Figure 5 shows a sample discussion from two students with learning losses (i.e., negative learning outcomes). In this chat log there were very few exchanges of authoritative statements, and this example is exemplary of the other authoritative exchanges through the rest of the discussion. Notice that many of the contributions are responses submitted to be evaluated, often denoted by question marks and coded as secondary knower statements. The two primary knower statements on lines 5 and 8 are either uncertain evaluations of the other student's authoritative contribution, or are very brief. Similar to Figure 4, we also see here a trend throughout the design activity where Student3 performed more primary knower statements than his partner. We performed analyses similar to (and including) the examples above in order to explore authoritativeness as a social factor affecting learning in collaborative settings.

	Name	Text	Authority
1	Tutor	...Remember the important relationship between Pmin and Environmental impact. Try to think about how much Pmin can be reduced without violating any constraints.	
2	Student3	so if pmin is about...110? Kpa	K2
3	Student4	is that above .85?	K2
4	Student4	quality	
5	Student3	probably not, i can't eyeball it	K1
6	Tutor	Right .	
7	Student4	125ish?	K2
8	Student3	k	K1
9	Student4	i think tutor's insulting my ability to read graphs	

Figure 5. A Sample of Discussion between Two Students with Poor Learning Gains, Coded for Authoritativeness.

Methods

The data we analyze in this paper was collected as part of a research study in which alternative forms of support for online learning were contrasted in a mechanical engineering course. 106 undergraduate students from a thermodynamics class at Carnegie Mellon University participated in the study by attending one of six lab sessions, in which time was strictly controlled. Students were given software training and practice (60 minutes), a pre-questionnaire and pre-test (10 minutes), the experimental manipulation (40 minutes), and then the post-test and post-questionnaire (15 minutes). The experimental design activity consisted of randomizing students to pairs, and then assigning each partner to design either an eco-friendly power plant or a power- proficient power plant. Opposing goals were used to encourage discussion and negotiation amongst partners. In all conditions, a tutor agent participated with the students in the chat in order to offer support. The entire lab session took place in a single computer lab, in which the researchers ensured that partners did not sit next to each other. The experimental manipulation took place during an online collaborative design discussion and consisted of modifying tutor behaviors only. In all other respects, the student experience in all conditions was the same. Results of the experimental manipulation are not discussed in this paper as they have been published separately (Ai et al., 2010).

Software

Students used Cyclepad (Kumar et al, 2007), a computer software simulator that allows users to implement thermodynamics design ideas via a graphical interface. Specifically, students must consider trade-offs between power output and environmental friendliness in designing a Rankine cycle, which is a type of heat engine. Essentially, students use Cyclepad to design simulated power plants.

Pairs of students used ConcertChat (Stahl, 2006), collaboration software enabling communication through a chat window (similar to instant messaging) and a whiteboard for sharing graphical information. ConcertChat has its origins in explicit referencing research, as discussed in Muehlpfordt & Wessner (2005) and Pfister & Muehlpfordt (2002).

Experimental Design

The experimental manipulation was a 3X3 between-subjects design. Each student pair was randomly assigned to one of nine conditions. The first independent variable contrasted 3 social conditions (High, Low, and None) where tutoring agents presented differing amounts of social behavior within the chat environment. Our dialogue agent exhibited three different types of socio-emotional behavior in the chat window: showing solidarity, tension release, and agreeing. The frequency of social behavior in our socially capable tutors was determined by a percentage of tutor turns that can be social prompts; specifically, the threshold parameter was 15% in the Low social tutor condition and 30% in the High social tutor. There was no social behavior in the Non-social condition. The task related behavior of the tutor was the same in all three of these conditions. Only the social aspects of the tutor's behavior changed.

For the second three-level independent variable, we designed 3 conditions in which the dialogue agent showed goal alignment either with the Green goal students, the Power goal students, or neither. This manipulation affected only the manner in which task related contributions were made in the conversation, but content was held constant. In this way, students could be in one of three different conditions in relation to the tutor agent, namely: Match (where the student's goal orientation condition matched the alignment of the tutor), Mismatch (where the student's condition is the opposite of the goal alignment exhibited by the tutor), or Neutral (where the tutor showed no bias). In all cases, the tutor presented the same task information. The only difference between conditions was the bias exhibited. For example, where the Green biased tutor might say "What is bad about increasing the heat input to the cycle is that it increases the heat rejected to the environment," the neutral tutor would say "Increasing heat input to the cycle increases the heat rejected to the environment."

Outcome Measures

As outcome measures, we examined:

- Learning gains between the Pre- and Post- test. 35 isomorphic multiple choice and short answer questions were used to test analytical and conceptual knowledge of Rankine cycles.
- The pre-questionnaire consisted of a brief scale for measuring mastery related beliefs (said to predict self-efficacy), and an alternate version of a self-efficacy question as shown in Figure 6. We combined the mean response of these four questions to form a self-efficacy scale.
- The post-questionnaire was a measure of students' perceptions of task success, and assessment of the quality of the interaction with their partner and with the agent.
- We also analyzed conversational behavior in the chat logs, with respect to their green and power biased statements.
- Conversational behavior was also analyzed with our Authoritativeness Framework.

Q10	I have always had a natural talent for engineering-related subjects.
Q11	I received good grades in my high school math classes.
Q12	I have always done well on science course assignments.
Q13	I am certain I will have completed the "Thermodynamics 1" course well.

Figure 6. Sample Pre-questionnaire Items Measuring Self-efficacy and Mastery Beliefs.

Results

Main effects of data from this study were originally reported and discussed in depth in Ai et al (2010), so here we will discuss new findings related to the self-efficacy pre-questionnaire and authoritativeness analysis. The main effects from the original analysis can be quickly summarized as follows: with the two independent variables (i.e., social manipulation and tutor goal map), the objective section of the pretest as covariate, and lab session as a random variable, there was a significant effect of Social Manipulation ($F(2,94) = 5.27$, $p < .01$) on learning where the Low Social condition was significantly better than the other two, with an effect size of .83 standard deviations in both cases. Other results related to post-questionnaire data and goal-biased conversational data are reported in Ai et al (2010).

Throughout this new analysis, "authoritative ratio" refers to the total number of primary knowing (i.e., K1) statements, over the total number of authoritativeness statements (i.e., K1+K2). Additionally, our results related to authoritativeness did not show statistically significant differences between conditions, and as such, we are not including tutor behavior in our analysis.

Learning Outcomes

The first dimension of our analysis involves examining learning outcomes and how various factors affected student learning. Posttest scores were regressed on the pretest and self-efficacy, but the relationship between self-efficacy and posttest was not significant. In short, individual self-efficacy does not seem to predict learning in this experiment. We can combine authoritativeness ratio and partner-ratio to predict group self-efficacy (the mean of both partner's self-efficacy), and we find a marginal effect of group self-efficacy on learning ($F(1, 103) = 3.29$, $p = 0.073$*).

It is important to note that group self-efficacy can be an important factor to consider. Wang & Lin (2007) found that individual student self-efficacy predicted collective efficacy, and collective efficacy predicted use of high-level cognitive skills in discussion, as well as group performance. In this study we are not currently looking at the usage of high-level cognitive skills in discussion, but we are curious about what other outcomes group self-efficacy may predict.

Looking at our other factor of interest, a linear regression analysis finds that authoritativeness ratio has a significant effect on learning ($F(1, 103) = 4.58$, $p = .0347$**), explaining 41% of the variance ($R^2 = .41$) or 3% of the variance beyond what is explained by the pretest. Where self-efficacy does not predict learning, authoritativeness appears to do so.

Self-Efficacy Beliefs and Authoritativeness

Since authoritativeness predicted performance and self-efficacy unexpectedly did not, it is not surprising that individual self-efficacy is not correlated with authoritativeness by itself. However, when we look at group self-efficacy we find a significant relationship ($F(1, 103) = 8.60$, $p = 0.0041$**) which explains 12% of the variance. So, despite authoritativeness and individual self-efficacy not being related on an individual level, when both partners' authoritativeness ratios are included in an analysis, authoritativeness can predict the group's self-efficacy, which may be more important in a collaborative setting.

Self-Efficacy Beliefs and Goal-Biased Conversation

In our originally published analysis, we investigated how our experimental manipulation affected the goal related bias displayed by students in their conversational behavior. For this analysis, we measure the bias of a student utterance towards the "green goal" or the "power goal" by applying a topic discovery model on our dialogues (Paul & Girju, 2009). Latent Dirichlet Allocation (LDA) models have been widely used to discover topics on large collections of unannotated data by modeling the word distributions represented in the data (Blei et al, 2003). We are using LDA models for the purpose of modeling how users are interacting with each other. For each utterance, we compute a score to represent to which degree the utterance displays a bias towards the green or power goals. When referring to the score for a student's bias towards his own goal, we will affix the word "self." So, if we want to know the average score towards the green goal, for a student who was assigned the green (i.e., environmentally friendly) power plant, then we will look at her Self-Average goal-biased conversation score.

Looking at partner's authoritativeness ratio along with one's own authoritativeness ratio, we find that partner's authoritativeness ratio predicts the Self-Average goal-bias, $F(1, 103) = 4.47$, $p = 0.0369$**, but it only accounts for 4% of the variance ($R^2 = .4$), so this relationship is not particularly informative. Nevertheless, it is interesting that we find evidence that suggests that speakers may respond to the authoritativeness with which their partner is positioned by exaggerating the distinction between his or her goal affiliation and his or her partner's. Further supporting the interpretation that a student's positioning is sensitive to the positioning of her or her partner, we find a negative correlation between the student's authoritativeness ratio and that of the partner, $R(106) = -.27$, $p = .0053$**.

Conclusion

We have seen with this set of data that authoritativeness predicts student learning from a collaborative activity better than self-efficacy. Additionally, both partners' authoritativeness predicts their group self-efficacy (even though individual authoritativeness does not predict individual self-efficacy). We also see a relationship between partner authoritativeness and how a student presents him or herself, both in terms of the extent to which topic affiliation is emphasized as well as personal authoritativeness. Whereas in studies of individual learning self-efficacy predicts learning, in this collaborative learning setting we find that there is a relationship between authoritativeness and self-efficacy, and authoritativeness and learning, but not self-efficacy and learning. These findings suggest that we cannot always rely on disposition research originating in individual settings to predict the same outcomes in collaborative learning settings. Other research on collaborative learning shows that these individual dispositions can predict other behaviors, such as method of conflict resolution, but our results suggest that authoritativeness may be a new disposition that is related to self-efficacy and that can predict learning in collaborative settings. While dispositions are typically considered to have personal causality, the authors believe authoritativeness to be influenced both by personal causes and situational causes to some extent.

This research reveals a considerable amount of potential for explaining collaborative learning through student dispositions and social behaviors. Future work will involve looking closer at the relationship between authoritativeness and other dispositions used to predict learning, such as achievement goal orientation. Additionally, knowing what authoritative combinations of partners produces the best learning gains could also be beneficial for the research community. It would also be necessary to explore how generalizable the effect of authoritativeness is, and if it is applicable to other domains or even learning tasks that are not project-based. Further work is also necessary to automate the process of assessing authoritativeness. If authoritativeness can be measured automatically via computer, it will have additional benefits beyond dispositional factors that are measured through self-report.

In conclusion, this paper shows that authoritativeness is an influential factor in discussion-based collaborative learning tasks that predicts learning and is related to a group's self-efficacy. Further research is required to refine the measurement of authoritativeness, and explore exactly how it is related to other dispositions. The end goal is to better understand the factors that influence learning in collaborative settings so that they may be automatically assessed in real world settings.

References

Ai, H., Kumar, R., Nagasunder, A., Rosé, C. P. (2010). Exploring the effectiveness of social capabilities and goal alignment in computer supported collaborative learning, *Intelligent Tutoring Systems conference (ITS 2010).*

Bandura, A. (1977). Self-Efficacy: Toward a unifying theory of behavioral change, *Psychological Review*, *84*, 191-215.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Blei, D. Ng, A., and Jordan, M. (2003). Latent dirichlet allocation. Journal of Machine Learning Research, 3.
Coutinho, S.A. & Neuman, G. (2008). A model of metacognition, achievement goal orientation, learning style and self-efficacy, *Learning Environments Research*, *11*, 131-151.
Darnon, C., Muller, D., Schrager, S. M., Pannuzzo, N., & Butera, F. (2006). Mastery and performance goals predict epistemic and relational conflict regulation. *Journal of Educational Psychology*, *98,* 766–776.
Dweck, C.S. (1986). Motivational processes affecting learning, *American Psychologist*, *41*,1040-1048.
Eccles, J.S. & A. Wigfield, (2002). Motivational beliefs, values, and goals, *Annual Review of Psychology*, *53*, 109-132.
Harackiewicz, J.M., Pintrich, P.R., Barron, K.E. & A. J. Elliot (2002), Revision of achievement goal theory: necessary and illuminating. *Journal of Educational Psychology*, *94*, 638-645.
Howley, I., Mayfield, E., & Rosé, C. (2011). Linguistic Analysis Methods for Studying Small Groups. *Invited Book Chapter to Appear in International Handbook of Collaborative Learning*.
Kumar, R., Rosé, C. P., Wang, Y. C., Joshi, M., and Robinson, A. (2007). Tutorial Dialogue as Adaptive Collaborative Learning Support. Proceedings of Artificial Intelligence in Education.
Martin, J. (1992). *English Text: System and structure*. Amsterdam: Benjamins.
Martin, J. (2000). Factoring out Exchange: Types of Structure. M Coulthard, J Cotterill and F Rock [Eds.] *Working with Dialogue*. Tubingen: Niemeyer. 19-40.
Martin, J. R., & Rose, D. (2003). *Working with Discourse: Meaning Beyond the Clause*. Continuum
Martin, J.R., Zappavigna, M., & Dwyer, P. (2008). Negotiating Shame: exchange and genre structure in youth justice conferencing. *European Systemic Functional Linguistics Conference*.
Mayfield, E. & Rosé, C. P. (2011). Recognizing Authority in Dialogue with an Integer Linear Programming Constrained Model, in *Proceedings of the Annual Meeting of the Association for Computational Linguistics.*
Muehlpfordt, M., & Wessner, M. (2005). Explicit Referencing in Chat Supports Collaborative Learning. In T. Koschmann, D. D. Suthers, and T.-W. Chan, editors, *Computer Supported Collaborative Learning 2005. The Next 10 Years!* (CSCL 2005), Mahwah, NJ, 2005. Lawrence Erlbaum Associates.
Newman, R. S. (1990). Children's help-seeking in the classroom: The role of motivational factors and attitudes. *Journal of Educational Psychology*, *82*(1), 71-80.
Nicholls, J.G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review*, 91, 328-346.
Paul, M., and Girju, R. (2009). Cross-Cultural Analysis of Blogs and Forums with Mixed-Collection Topic Models Export. In P*roceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*.
Pfister, H.R., & Muehlpfordt, M. (2002). Supporting Discourse in a Synchronous Learning Environment: The Learning Protocol Approach. In G. Stahl, editor, *Proceedings of CSCL 2002*, pp 581–589, Hillsdale, 2002. Erlbaum.
Piaget, J. (1985). The equilibrium *of cognitive structures: the central problem of intellectual development*, Chicago University Press.
Ryan, A. M., & Pintrich, P. R. (1997). "Should I ask for help?" The role of motivation and attitudes in adolescents' help seeking in math class. *Journal of Educational Psychology*, *89*(2), 329-341).
Sheldon, K. M., & Kasser, T. (1995). Coherence and congruence: Two aspects of personality integration. *Journal of Personality and Social Psycholog*y, *68*(3), 531-543.
Sheldon, K. M., Ryan, R. M., Deci, E. L., & Kasser, T. (2004). The independent effects of goal contents and motives on well-being: It's both what you pursue and why you pursue it. *Personality and Social Psychology Bulletin*, 30, 475-486.
Stahl, G. (2006). Analyzing and designing the group cognitive experience. *International Journal of Cooperative Information Systems (IJCIS)*.
Veel, R. (1999). Language, Knowledge, and Authority in School Mathematics, in Francis Christie (Ed.) *Pedagogy and the Shaping of Consciousness: Linguistics and Social Processes*, Continuum.
Wang, S. & Lin, S. (2007). The effects of group composition of self-efficacy and collective efficacy on computer-supported collaborative learning, *Computers in Human Behavior*, *23*(5), 2256-2268.
Zimmerman, B. J. (1999). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology, 25*, 82-91.

Acknowledgments

This work was supported by the Pittsburgh Science of Learning Center (#SBE 0836012) and a Graduate Training Grant from the Department of Education (#R305B040063).

Fostering Social Navigation and Elaboration of Controversial Topics with Preference-Inconsistent Recommendations

Christina Schwind, Jürgen Buder, Friedrich W. Hesse, Knowledge Media Research Center, Konrad-Adenauer-Str. 40, 72072 Tübingen, Germany
Email: c.schwind@iwm-kmrc.de, j.buder@iwm-kmrc.de, f.hesse@iwm-kmrc.de

Abstract: Critical thinking requires knowledge about the diversity of viewpoints on controversial issues. However, the diversity of perspectives often remains unexploited: Learners prefer preference-consistent over preference-inconsistent information, a phenomenon called confirmation bias. Two lab experiments were designed to test whether technologies such as recommender systems can be used to overcome this bias. The role of preference-inconsistent recommendations was explored by comparing their influence to a condition with preference-consistent recommendations and to a control condition without recommendations. In Study 1, preference-inconsistent recommendations led to a reduction of confirmation bias and to an attenuation of preferences. In Study 2, we found that preference-inconsistent recommendations stimulated balanced recall and divergent thinking. Together these studies showed that preference-inconsistent recommendations can foster social navigation and elaboration. In conclusion, future research and practical implications are discussed.

Introduction

Controversies and debates on political or health related issues can be characterized as ongoing deliberations among stakeholders with widely diverse perspectives. On the basis of these perspectives, individuals are able to form an opinion and to back up this opinion with arguments. Thus, learning is a central part of opinion formation and can help the individual to make an informed decision. The WWW is a perfect backdrop for this kind of learning as a multitude of opinions is publicly available: Whether it is through information portals on controversial issues, or through the exchange on forums, social networks, or other channels, information spaces can be characterized by an accumulation of a vast amount of differing opinions.

Why Information Search May Be Biased

However, the availability of different opinions remains unexploited: When learners only have a vague idea about a controversial issue, they inform themselves about the facts, arguments, or explanations. When doing so, they often fail to take dissenting information into account which is referred to as selective exposure (Knobloch-Westerwick & Meng, 2010) or confirmation bias (Jonas, Schulz-Hardt, Frey, & Thelen, 2001). Festinger's (1957) dissonance theory provides the basis for this effect: Individuals try to avoid dissonance and therefore prefer information which supports their own position. Consequently, knowledge acquisition is typically biased in favor of the learner's previously held expectations. This biased knowledge acquisition leads to learners adhering to their position, even though considering others and their perspectives is a necessary first step for critical thinking (Stanovich & West, 1997). Therefore confirmation bias is likely to be a hindrance to critical thinking (West, Toplak, & Stanovich, 2008).

Why We Can Use Technology to Overcome Confirmation Bias

The question arises if and how we can use technology to overcome this bias. One possible solution is to emphasize preference-inconsistent information by the use of recommender systems, as they are optimized for making specific information salient. Recommender systems can be classified as a collaborative technology for filtering information: Users express social judgments explicitly by rating stimuli or implicitly by their navigational behavior. Out of these social judgments, each user receives personalized recommendations that are matched to the user's profile stored by the system (Konstan & Riedl, 2003). Community-generated recommendations may act as an information signpost influencing individual navigation and item selection. If this is the case, it is called social navigation (Dieberger, Dourish, Höök, Resnick, & Wexelblat, 2000).

In many ways, the principles of recommender systems are very much in line with principles that play an important role in the learning sciences. First, constructivist approaches often stress the importance of moving from teacher-centered to learner-centered education (Bransford, Brown, & Cooking, 1999). Recommender systems fit into this philosophy, as the recommendations do originate from the collective of other peers. Second, aptitude-treatment interaction expresses the idea of adapting information to the needs and abilities of learners (Cronbach & Snow, 1977). This is addressed by personalized recommender systems in which recommendations are specifically tailored to the learner. Finally, learners need some form of scaffolding (Vygotsky, 1978). Recommender systems provide information to learners on how to allocate their cognitive capacity. They can help to identify the resources that really matter and thereby create a zone of proximal development.

However, some features of recommender systems can be detrimental to learning. Classical recommender systems provide mainly preference-consistent recommendations, as this method appears to be promising in fields of taste and consumption. But in doing so, classical recommender systems violate the educational claim of multiperspectivity and informational diversity (De Wit & Greer, 2008; Spiro & Jehng, 1990). Therefore it would seem to be useful to recommend preference-*inconsistent* arguments for educational purposes.

Why Preference-inconsistent Recommendations May Help

Educational literature states that one of the learning goals is to think decontextualized without the biasing influence of prior preferences and opinions. Learners should elaborate on controversial topics and evaluate arguments regardless of whether they tend towards the advantages or disadvantages of a specific topic. This form of unbiased reasoning is referred to as critical thinking (Stanovich, & West, 1997). Facing preference-inconsistency or disagreement is considered to stimulate critical thinking (e.g., Buchs, Butera, Mugny, & Darnon, 2004; Johnson & Johnson, 2009). The basis of this approach is Piaget's (1950) theorizing that internal conflicts are necessary to stimulate cognitive development. The uncertainty which results from conflicting situations leads to epistemic curiosity (Berlyne, 1960), which is a main trigger for further information search and for an interest in an understanding of other perspectives. Doise and Mugny (1984) stressed that constructive conflicts are best initiated by the direct confrontation with the opinion of others. Further, social psychology hints at other potentials of conflicts, for instance in the literature on the influence of minorities (Nemeth, 2003). Participants confronted with authentic dissent from a minority show increased searching for information, deep elaboration, creative problem solutions, and thus divergent thinking (Nemeth, Connell, Rogers, & Brown, 2001; Nemeth, & Rogers, 1996; Nemeth, & Wachtler, 1983). From research on the influence of minorities, it is known that these effects can lead to informational influence and to changes in attitudes (Wood, Lundgren, Ouellette, Busceme, & Blackstone, 1994).

The question remains whether conflicting information must originate from direct interaction with another person or whether a preference-inconsistent recommendation originating from a computer system may also stimulate conflict and foster deeper elaboration. One indication that computer-based recommendations can be efficient substitutes relates to the notion that human social categories are ascribed to recommender systems (McNee, Riedl, & Konstan, 2006). However, empirical clarification is needed to investigate whether or not the "recommender personality" is sufficient to stimulate socio-cognitive conflict.

The present research attempted to clarify whether preference-inconsistent recommendations can stimulate conflict and thus affect social navigation and elaboration. We did not address the technical specifications of a recommender system, but rather explored the psychological impact of recommendations. In two lab experiments, we manipulated the type of social navigation tool by providing recommendations that were either consistent or inconsistent to the participants' prior preferences. These two conditions were compared to a control condition without any recommendation.

Hypotheses

Conflict can lead to epistemic curiosity and therefore trigger further information search and higher interest in understanding an opposite perspective (Buchs et al., 2004). Preference-inconsistent recommendations are conceptualized as stimulating conflict and therefore motivate learners to search for more inconsistent information resources.

Hypothesis 1. Therefore, it was hypothesized that participants will show natural confirmation bias when no recommendation is given. Preference-consistent recommendations will enhance confirmation bias, whereas preference-inconsistent recommendations will reduce confirmation bias. (Study 1 & Study 2)

Hypothesis 2. It was predicted that the difference in confirmation bias will have its equivalent in the adaptation of subjects' preferences: Participants' preferences will be strengthened in the post-preference when no recommendation is given. Preference-consistent recommendations will also strengthen preferences, whereas preference-inconsistent recommendations will weaken preferences. (Study 1)

Hypothesis 3. It was expected that participants confronted with preference-inconsistent recommendations are more likely to experience dissent. Therefore they will (*3a*) recall more arguments and (*3b*) show a more balanced recall than participants confronted with preference-consistent or no recommendation. (Study 2)

Hypothesis 4. Based on the literature on minority dissent, it was also expected that participants confronted with preference-inconsistent recommendations will *(4a)* generate more arguments in general as well as *(4b)* more novel arguments than participants confronted with preference-consistent or no recommendation. (Study 2)

Study 1

The first study was conducted as an online experiment and investigated the impact of recommendations on information selection and preference adaptation.

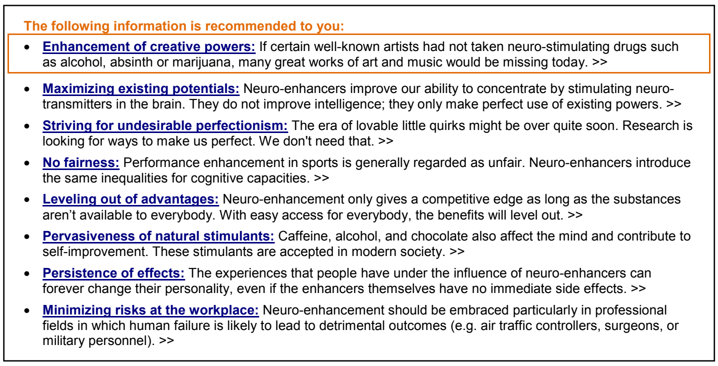
Method

Participants

One hundred twenty-five subjects participated in the experiment. They were recruited via an academic e-mail-list and compensated by the participation in a lottery. Based on manipulation check data, four subjects were excluded from the analysis. Ultimately, data of 121 subjects ($M = 25.28$, $SD = 6.05$; 92 female) were included in the analysis. In order to prevent high levels of prior knowledge, students of medicine or pharmacy were excluded from participation. The majority of participants (90.8 %) judged their prior knowledge of the subject matter to be very poor, poor, or average.

Materials

The materials of the online study comprised a number of screens. One screen contained a short introductory text about the controversial topic of neuro-enhancement, referring to the facilitation of cognitive abilities through training or through medication. The main screen of the experiment consisted of a bogus list of Web search results. The list was composed of eight written arguments, four of them supporting and four of them opposing neuro-enhancement. The arguments consisted of a headline followed by two explanatory sentences. An example of an argument supporting neuro-enhancement is: "Minimizing risks at the workplace: Neuro-enhancement should be embraced particularly in professional fields in which human failure is likely to lead to detrimental outcomes (e.g. air traffic controllers, surgeons, or military personnel)." An example of an argument opposing neuro-enhancement is: "Striving for undesirable perfectionism: The era of lovable little quirks might be over quite soon. Research is looking for ways to make us perfect. We don't need that." The arguments were tested in an online survey ($N = 48$) to balance them for credibility, persuasiveness, comprehensibility, originality, and strength. Although the headlines were marked as hyperlinks, participants could not access additional information. One of the arguments was highlighted by an orange-colored frame surrounding the text. This argument represented the recommendation. The caption above the frame stated: "The following information is recommended to you." (see Figure 1). The order of the arguments and the serial position of the recommendation were randomized across participants in order to minimize content and order effects.



The following information is recommended to you:

- Enhancement of creative powers: If certain well-known artists had not taken neuro-stimulating drugs such as alcohol, absinth or marijuana, many great works of art and music would be missing today. >>

- Maximizing existing potentials: Neuro-enhancers improve our ability to concentrate by stimulating neuro-transmitters in the brain. They do not improve intelligence; they only make perfect use of existing powers. >>
- Striving for undesirable perfectionism: The era of lovable little quirks might be over quite soon. Research is looking for ways to make us perfect. We don't need that. >>
- No fairness: Performance enhancement in sports is generally regarded as unfair. Neuro-enhancers introduce the same inequalities for cognitive capacities. >>
- Leveling out of advantages: Neuro-enhancement only gives a competitive edge as long as the substances aren't available to everybody. With easy access for everybody, the benefits will level out. >>
- Pervasiveness of natural stimulants: Caffeine, alcohol, and chocolate also affect the mind and contribute to self-improvement. These stimulants are accepted in modern society. >>
- Persistence of effects: The experiences that people have under the influence of neuro-enhancers can forever change their personality, even if the enhancers themselves have no immediate side effects. >>
- Minimizing risks at the workplace: Neuro-enhancement should be embraced particularly in professional fields in which human failure is likely to lead to detrimental outcomes (e.g. air traffic controllers, surgeons, or military personnel). >>

Figure 1. Screenshot of the Experimental Conditions with Recommendation (Translated from German).

Design

The experiment is based on a one factorial design with three conditions (control condition vs. preference-consistent vs. preference-inconsistent). In the control condition (CC), no argument was recommended to the participants. In the first experimental condition (EC_{con}), an argument was recommended that was consistent with the participant's pre-preference for or against neuro-enhancement. In the second experimental condition (EC_{inc}), an argument was recommended that was inconsistent with the participant's preference. The preference data showed that 70.2 % of the participants were against neuro-enhancement. This preference ratio was equally distributed over the three conditions ($F(2,118) < 1$, $p = .727$, $\eta^2 = .00$).

Procedure

The scenario was adapted from Jonas et al. (2001) and structured into three phases: First, subjects read the introductory passage on neuro-enhancement, and subsequently indicated their pre-preference on neuro-enhancement. In the second phase, the list of eight arguments was presented with four arguments in favor of and four arguments against neuro-enhancement. This phase varied depending on the condition, mimicking the personalization part of recommender systems. In the control condition, no recommendation was given. In the consistent experimental condition, subjects received a preference-consistent recommendation, and in the inconsistent experimental condition, subjects received a preference-inconsistent recommendation. Subjects were asked to select exactly one of the eight arguments that they would like to read more about by clicking on an adjacent box. In the third phase, participants were asked to indicate their post-preference. The preference adaptation scales contained one item on a 6-point bipolar continuum. The preference indication was represented by a word pair ranging from opposing (-2.5) to endorsing (+2.5) neuro-enhancement. Afterwards, participants were asked to fill out two items for the manipulation check and to provide demographic details (age, gender).

Measures

In the first study, we focused on information selection and preference adaptation.

Manipulation check. Manipulation check consisted of two questions. Participants had to decide if the recommended argument was for or against neuro-enhancement on a dichotomous item and they were asked to indicate on a 5-point Likert scale whether the recommended argument matched their own position *absolutely* (+2) or *not at all* (-2).

Information selection. Information selection was measured for testing the impact of the recommendations on confirmation bias. The measurement expressed the likelihood of selecting a preference-consistent over a preference-inconsistent argument.

Preference adaptation. The adaptation of preferences was calculated as the difference between the absolute value of the post-preference and the absolute value of the pre-preference. In this way, it was possible to differentiate whether subjects' preferences were weakened or strengthened.

Results

Information selection was tested with χ^2–tests, whereas preference adaptation was analyzed using analysis of variance (ANOVA).

Manipulation check. The first question asked whether the recommended argument was for or against neuro-enhancement. For the second question, EC_{con} subjects were excluded if they indicated that the recommended argument did not match their pre-preference (i.e. a value less than or equal to zero). Contrary, EC_{inc} subjects were excluded if they indicated that the recommended argument did match their pre-preference (i.e. a value greater than or equal to zero). Participants were included in the analysis only if the answers to both questions were correct. Following this rule, the analysis comprised 40 CC subjects, 41 EC_{con} subjects, and 40 EC_{inc} subjects.

Information selection. In Hypothesis 1, a confirmation bias was predicted in the control condition, an enhancement of confirmation bias in the preference-consistent condition, and a reduction of confirmation bias in the preference-inconsistent condition. To test this prediction, we computed an overall χ^2–test with the factors condition (CC vs. EC_{con} vs. EC_{inc}) and information selection (consistent vs. inconsistent). The analysis revealed a marginally significant effect, $\chi^2_{0,95\ (2)} = 3.51$, $p = .087$, $d = 0.35$. By testing the three conditions separately, a significant effect occurred for CC ($\chi^2_{0,95\ (1)} = 8.10$, $p = .002$, $d = 1.01$) as well as for EC_{con} ($\chi^2_{0,95\ (1)} = 10.76$, $p < .001$, $d = 1.19$). For EC_{inc}, no confirmation bias could be detected ($\chi^2_{0,95\ (1)} < 1$, $p = .343$, $d = 0.30$); see Figure 2. This effect occurred although the acceptance rates of the recommendations (the likelihood that the recommended argument was selected) differed between the two experimental conditions: preference-consistent recommendations were accepted by 42 % of subjects, whereas preference-inconsistent recommendations were accepted by 20 % of subjects.

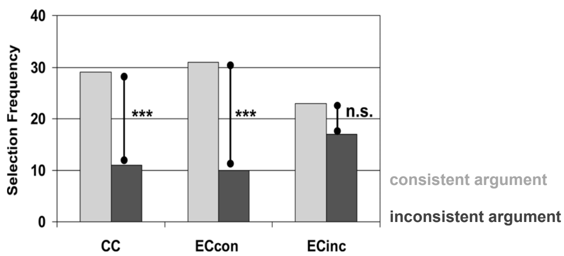


Figure 2. Selection Frequencies in the Three Conditions (Study 1).

Preference adaptation. We expected a weakening of post-preference vs. pre-preference in the condition with preference-inconsistent recommendation and a strengthening in the other two conditions (Hypothesis 2). The analysis of the adaptation of participants' preferences for the three conditions yielded the expected main effect, $F(2,118) = 3.66$, $p = .029$, $\eta^2 = .06$. Participants in the condition with preference-inconsistent recommendation indicated a more moderate view; their preference became weaker (M = -0.30, $SD = 0.65$).

Participants in the control condition ($M = 0.00$, $SD = 0.59$) as well as those in the condition with preference-consistent recommendations ($M = 0.05$, $SD = 0.64$) did not change their preference. Pairwise comparisons using simple contrasts revealed that EC_{inc} subjects had stronger preference weakening compared to CC subjects ($p = .014$), while there was no effect for EC_{con} subjects compared to CC subjects ($p = .720$).

Discussion

Study 1 provides first evidence for the impact of recommendations on information selection and preference adaptation. As expected in Hypothesis 1, CC subjects showed confirmation bias by choosing preference-consistent over preference-inconsistent arguments. In EC_{con}, the confirmation bias was also present; however, we could not find the expected enhancement of the bias. In EC_{inc}, the expected reduction of confirmation bias occurred. These effects in information selection had their equivalent in preference adaptation (Hypothesis 2): CC subjects as well as EC_{con} subjects did not adapt their preferences. However, EC_{inc} led to a more moderate view of the controversial topic of neuro-enhancement. Based on Study 1, it is not clear whether or not these changes in information selection and preference adaptation were accompanied by deeper elaboration. In order to measure the impact of recommendations on elaboration, we conducted the follow-up Study 2.

Study 2

The second study was conducted as a lab experiment. Two goals were pursued in this study: First, we wanted to replicate our findings concerning the impact of recommendations on information selection. Second, we investigated the impact of recommendations on free recall and on the generation of opinion statements.

Method

Participants

One hundred one students from a German university participated in the experiment. They were recruited via an academic e-mail-list and compensated by either payment or course credit. Twelve subjects were excluded from the analysis either because of failing the experimental manipulation check or because of their medical related field of study. Therefore, data of 89 subjects ($M = 23.15$, $SD = 3.11$; 62 female) were included in the analysis. The majority of participants (84.3 %) judged their prior knowledge about the subject matter to be very poor, poor, or average.

Material

The simulated website with search results used in the first study as well as the application domain of neuro-enhancement remained the same.

Design

The experiment again was based on a one factorial design with the three conditions (CC vs. EC_{con} vs. EC_{inc}). The pre-preference data showed that 82 % of the participants were against neuro-enhancement. This preference ratio was equally distributed over the three conditions ($F(2,86) < 1$, $p = .781$, $\eta^2 = .00$).

Procedure

The procedure remained the same as in Study 1 except for the third phase including measuring the dependent variables. As in the first study, subjects were asked to select exactly one argument that they would like to read more about by clicking on an adjacent box. In the third phase, participants were asked for free recall of the eight arguments on the simulated website. This free recall test was followed by an opinion statement: participants were asked to state their opinion and to justify it by writing an essay. They were explicitly instructed that they did not have to use all information they could remember (in contrast to the free recall task) and that they could use new and self-created arguments respectively. Afterwards, participants were asked to fill out two items for the manipulation check and to provide demographic details (age, gender).

Measures

In the second study, we focused on free recall and on the generation of opinion statements. The measures for manipulation check and information selection remained the same as in the first study (see above). Therefore we will describe here only the new measures for elaboration, namely, free recall and generation of opinion statements.

Free recall. First, we counted the overall number of arguments recalled. Second, we calculated a balancing index [($arguments_{consistent}$ recalled/$arguments_{inconsistent}$ recalled) – 1] to find out *which* arguments were recalled. A score around zero indicated a balanced recall which means that participants recalled an equal number of preference-consistent and preference-inconsistent arguments. A score above zero indicated a biased recall for preference-consistent arguments, and a score below zero indicated a biased recall in favor of preference-inconsistent arguments.

Opinion statement. For opinion statement, the overall number of arguments generated per participant was counted. Further, we were interested in the percentage of novel arguments generated in the essay. For the analyses, a coding schema suggested by Cacioppo, Harkins, and Petty (1981) was used. Accordingly, arguments were coded as 0 = "external" when they were part of the arguments on the search result website, or as 1 = "internal" when the arguments were not mentioned before and thus generated by the participants. All 89 essays were coded by two raters. The interrater reliability Cohen's Kappa was $K = .88$ which indicates an almost perfect strength of agreement.

Results

Information selection was tested with χ^2–tests, whereas elaboration measures were analyzed using ANOVAs.

Manipulation check. Following a criterion set for the first study (inclusion into the analysis only if both answers were correct), the analysis comprised 30 CC subjects, 29 EC_{con} subjects, and 30 EC_{inc} subjects.

Information selection. In this study, we attempted to replicate the findings of the first experiment concerning information selection (Hypothesis 1). Therefore, we computed an overall χ^2–test with the factors condition (CC vs. EC_{con} vs. EC_{inc}) and information selection (consistent vs. inconsistent). The analysis yielded the expected effect, $\chi^2_{0,95\,(2)} = 3.99$, $p = .068$, $d = 0.43$. We also found the same pattern as in the first experiment for the three conditions separately: a significant effect occurred for CC ($\chi^2_{0,95\,(1)} = 6.53$, $p = .011$, $d = 1.06$) as well as for EC_{con} ($\chi^2_{0,95\,(1)} = 4.17$, $p = .041$, $d = 0.82$). For EC_{inc}, no such effect occurred ($\chi^2_{0,95\,(1)} < 1$, $p = .999$, $d = 0.00$); see Figure 3. This effect occurred although the acceptance rates of the recommendations differed between the two experimental conditions: Preference-consistent recommendations were selected for further information from 24 % of subjects, whereas preference-inconsistent recommendations were selected for further information from 10 % of subjects.

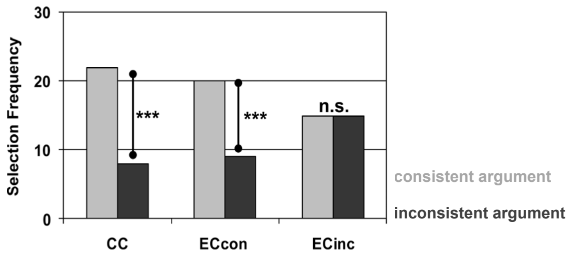


Figure 3. Selection Frequencies in the Three Conditions (Study 2).

It was predicted that information selection is related to information processing such that less confirmation bias results in deeper elaboration. Therefore, we expected free recall (Hypothesis 3a and 3b) and opinion statement (Hypotheses 4a and 4b) to be best in EC_{inc}. In order to test the hypotheses, one-factorial ANOVAs with condition (CC vs. EC_{con} vs. EC_{inc}) as the independent variable and different elaboration measures as dependent variables were conducted.

Free recall. The analysis of number of arguments recalled revealed no difference between the three conditions ($F(2,86) < 1$, $p = .623$, $\eta^2 = .00$). CC subjects recalled $M = 4.33$ ($SD = 1.09$) arguments, EC_{con} subjects $M = 4.00$ ($SD = 1.75$) arguments, and EC_{inc} subjects $M = 4.03$ ($SD = 1.45$) arguments. The analysis of the recall balancing index yielded the expected highly significant main effect, $F(2,81) = 5.03$, $p = .009$, $\eta^2 = .04$: Subjects in the control condition ($M = 0.61$, $SD = 1.07$) as well as those in the condition with consistent recommendation ($M = 0.26$, $SD = 0.64$) recalled the arguments biased towards preference-consistency, whereas subjects in the condition with inconsistent recommendation showed a balanced recall ($M = -0.06$, $SD = 0.67$). Pairwise comparisons using simple contrasts revealed that EC_{inc} subjects showed stronger recall balancing compared to CC subjects ($p = .002$), while there was no significant effect for EC_{con} subjects compared to CC subjects ($p = .128$).

Opinion statement. The analysis of number of arguments generated revealed a significant effect between the three conditions ($F(2,86) = 3.13$, $p = .049$, $\eta^2 = .01$). CC subjects generated $M = 2.97$ ($SD = 1.19$) arguments, EC_{con} subjects $M = 2.97$ ($SD = 1.40$) arguments, and EC_{inc} subjects $M = 3.77$ ($SD = 1.66$) arguments. The analysis of the source of arguments yielded the expected highly significant main effect, $F(2,86) = 7.11$, $p = .001$, $\eta^2 = .06$: Participants in the control condition ($M = 1.10$, $SD = 0.92$) as well as those in the condition with consistent recommendation ($M = 1.34$, $SD = 1.17$) showed less internal argument generation compared to the condition with inconsistent recommendation ($M = 2.30$, $SD = 1.69$). Pairwise comparisons using simple contrasts revealed that EC_{inc} subjects showed stronger internal argument generation compared to CC subjects ($p = .001$), while there was no significant effect for EC_{con} subjects compared to CC subjects ($p = .472$).

Discussion

In Study 2, it was possible to replicate our findings for the impact of a recommendation on confirmation bias and to extend them in several ways. In line with Hypothesis 1, both studies demonstrated that CC subjects showed a confirmation bias. Confirmation bias remained stable for EC_{con} subjects and was reduced or even prevented for EC_{inc} subjects. It is worth noting that both studies revealed the same effect pattern concerning information selection, although the data collection (online and lab) as well as subjects' compensation (lottery and payment) differed. As the effect occurred in both studies, albeit in different contextual settings, it appears to be a robust finding. Study 2 found that EC_{inc} stimulated balanced recall, whereas CC and EC_{con} resulted in biased recall towards preference-consistent arguments (Hypothesis 3b). However, the overall number of arguments recalled revealed no difference between the three conditions (Hypothesis 3a). In addition, Study 2 demonstrated that EC_{inc} stimulated increased argument generation compared to CC and EC_{con} (Hypothesis 4a). Furthermore, EC_{inc} subjects generated more novel arguments compared to CC and EC_{con} subjects (Hypothesis 4b). This internal argument generation can be interpreted as divergent thinking processes stimulated by EC_{inc}.

General Discussion

The current research demonstrated that preference-inconsistent recommendations can foster social navigation and elaboration of controversial topics. Both studies showed that preference-inconsistent recommendations have an effect on information selection and thus can help to overcome confirmation bias. This effect was not associated with a high acceptance rate of preference-inconsistent recommendations: Participants did not select the recommended argument; they selected a different argument from the recommended perspective instead. This rejection could have resulted from psychological reactance (Miron & Brehm, 2006), as reactance is known to stimulate resistance when people perceive their freedom of choice to be restricted. Alternatively, the design of the recommendation used in the studies could have led to banner blindness (Benway & Lane, 1998), a phenomenon in which emphasized information is ignored. Further research is needed to ascertain which factors contribute to the low acceptance rate of recommendations and how an enhancement could be achieved. Furthermore, we could not find empirical evidence for our hypothesis that classical, preference-consistent recommendations lead to an enhanced confirmation bias. Therefore, the implementation of recommender systems for learning does not appear to have debilitating effects on informational diversity above natural confirmation bias. Consequently, it might be worthwhile to strive for wider application of recommender systems in educational contexts.

This claim is supported by our findings on further points. Study 1 demonstrated that preference-inconsistent recommendations weaken participants' preferences; participants formed a more moderate view on the controversial topic. Study 2 demonstrated the impact of preference-inconsistent recommendation on elaboration. Participants showed a balanced recall by remembering arguments from both perspectives and divergent thinking by generating more novel arguments. This difference between preference-consistent vs. preference-inconsistent recommendations might result from the difference in the perceived recommender personality. Taking the research of minority dissent into account, participants might have assumed that preference-inconsistent recommendations stemmed from a minority and thus leading to divergent thinking. Further studies should include an investigation of the perceived "humanization" of a recommender system.

Although our studies were conducted in a non-collaborative context, we believe that they have implications for CSCL. For instance, exposure to preference-inconsistent opinions is regarded as an important element of collaborative learning. Our findings implicate that exposure to dissenting information might not be enough to stimulate unbiased reasoning and critical thinking. This can be concluded from the fact that in all three conditions preference-inconsistent arguments were available to the learners. However, only when a preference-inconsistent argument was made salient, did the corresponding opinion have the persuasive power to influence navigation and elaboration. Whether visual salience suffices or whether learners attribute some authority to a recommendation requires further empirical investigation. However, it appears evident that without computer-supported salience, preference-inconsistent arguments can hardly unfold their beneficial effects.

Another relevant dimension for further research is the difference between debate and controversy (Johnson & Johnson, 1979). Preference-consistent recommendations highlight arguments which match the learners' preferences. Therefore consistent recommendations might trigger a debate mode of processing and lead to differentiation processes. On the other hand, preference-inconsistent recommendations highlight arguments which do not match the learners' preferences, but since they are denoted as a recommendation, they might be viewed more positively. This might trigger a controversy mode of processing and lead to integration processes.

This paper provides experimental evidence that in opinion formation contexts learners will only accept a challenge when the challenging information is recommended. In this regard, it can be said that preference-inconsistent recommender systems create a zone of proximal development. If properly implemented, recommender systems have the potential to become powerful tools supporting learners on their way to becoming critical thinkers and informed decision makers.

References

Benway, J. P., & Lane, D. M. (1998). Banner blindness: the irony of attention grabbing on the World Wide Web. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*. Chicago, IL.

Berlyne, D. E. (1960). *Conflict, arousal, and curiosity*. New York: McGraw-Hill.

Bransford, J. D., Brown, A., & Cooking, R. (Eds.) (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Buchs, C., Butera, F., Mugny, G., & Darnon, C. (2004). Conflict elaborations and cognitive outcomes. *Theory into Practice*, *43*(1), 23-30.

Cacioppo, J. T., Harkins, S. G., & Petty, R. E. (1981). The nature of attitudes and cognitive responses and their relationships to behavior. In R. E. Petty, T. M. Ostrom, & T. C. Brock (Eds.), *Cognitive responses in persuasion* (pp. 217-235). Hillsdale, NJ: Lawrence Erlbaum.

Cronbach, L., & Snow, R. (1977). *Aptitudes and instructional methods: A handbook of research on interactions.* New York: Irvington Publishers.

Dennis, A. R., & Valacich, J. S. (1999). Rethinking media richness: Towards a theory of media synchronicity. *Proceedings of the 32nd Hawaii International Conference on System Sciences*. Los Alamitos, CA.

De Wit, F. R. C., & Greer, L. L. (2008). The black-box deciphered: A meta-analysis of team diversity, conflict, and team performance. *Academy of Management Best Paper Proceedings*. Anaheim, CA.

Dieberger, A., Dourish, P., Höök, K., Resnick, P., & Wexelblat, A. (2000). Social navigation: Techniques for building more usable systems. *interactions*, *7*(6), 36-45.

Doise, W., & Mugny, G. (1984). *The social development of the intellect*. Oxford: Pergamon Press.

Festinger, L. (1957). *A theory of cognitive dissonance*. New York: Row Peterson.

Johnson, D. W., & Johnson, R. T. (1979). Conflict in the classroom: Controversy and learning. *Review of Educational Research*, *49*(1), 51-69.

Johnson, D. W., & Johnson, R. T. (2009). Energizing learning: The instructional power of conflict. *Educational Researcher*, *38*(1), 37-51.

Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology*, *80*(4), 557-571.

Knobloch-Westerwick, S., & Meng, J. (2009). Looking the other way: Selective exposure to attitude-consistent and counterattitudinal political information. *Communication Research*, *36*(3), 426-448.

Konstan, J. A., & Riedl, J. (2003). Collaborative filtering: Supporting social navigation in large, crowded infospaces. In K. Höök, D. Benyon, A. J. Munro (Eds.), *Designing information spaces: The social navigation approach* (pp. 43-81). London: Springer.

McNee, S. M., Riedl, J., & Konstan, J. A. (2006). Making recommendations better: An analytic model for human-recommender interaction. *Proceedings of the ACM Conference on Human Factors in Computing Systems*. Montreal, Canada.

Miron, A. M., & Brehm, J. W. (2006). Reactance theory – 40 years later. *Zeitschrift für Sozialpsychologie*, *37*(1), 9-18.

Nemeth, C. J. (2003). Minority dissent and its 'hidden' benefits. *New Review of Social Psychology*, *2*, 11-21.

Nemeth, C. J., Connell, J. B., Rogers, J. D., & Brown, K. S. (2001). Improving decision making by means of dissent. *Journal of Applied Social Psychology*, *31*(1), 48-58.

Nemeth, C. J., & Rogers, J. (1996). Dissent and the search for information. *British Journal of Social Psychology*, *35*, 67-76.

Nemeth, C. J., & Wachtler, J. (1983). Creative problem solving as a result of majority vs minority influence. *European Journal of Social Psychology*, *13*(1), 45-55.

Piaget, J. (1950). *The psychology of intelligence*. London: Routledge & Kegan Paul.

Spiro, R. J., & Jehng, J. C. (1990). Cognitive flexibility and hypertext: Theory and technology for the nonlinear and multidimensional traversal of complex subject matter. In D. Nix, & R. J. Spiro (Eds.), *Cognition, education, and multimedia: Exploring ideas in high technology* (pp. 163-205). Hillsdale, NJ: Lawrence Erlbaum Associates.

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology*, *89*(2), 342-357.

West, R. F., Toplak, M. E., & Stanovich, K. E. (2008). Heuristics and biases as measures of critical thinking: Associations with cognitive ability and thinking dispositions. *Journal of Educational Psychology*, *100*(4), 930-941.

Wood, W., Lundgren, S., Ouellette, J. A., Busceme, S., & Blackstone, T. (1994). Minority influence: a meta-analytic review of social influence processes. *Psychological Bulletin*, *115*(3), 323-345.

Vygotsky, L. S. (1978). In M. Cole, V. John-Steiner, S. Scribner, & E. Souberman (Eds. & Trans.), *Mind in society: the development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Adaptable Scripting in Computer-Supported Collaborative Learning to Foster Knowledge and Skill Acquisition

Xinghua Wang, Ingo Kollar, Karsten Stegmann, Frank Fischer, Ludwig-Maximilians-University of Munich, Leopoldstr. 13, 80802 Munich (Germany)
Email: wang.xinghua@psy.lmu.de, ingo.kollar@psy.lmu.de, karsten.stegmann@psy.lmu.de, frank.fischer@psy.lmu.de

Abstract: Collaboration scripts have repeatedly been implemented in CSCL with the aim of improving individual acquisition of domain-specific knowledge and domain-general skills. However, it is a delicate compromise between taking the "freedom from" (i.e., constraining the learners' efforts to self-regulate their learning) and providing the "freedom to" learners (i.e., enabling them to collaborate on a higher level) when implementing CSCL scripts. The study reported here aimed to answer the question whether adaptable scripting would be an effective approach to realize flexibility in scripted CSCL. 54 university students participated in this study which compared three experimental conditions (unscripted CSCL, CSCL with a non-adaptable script, CSCL with an adaptable script). Results showed that the adaptable script improved individual knowledge and skill acquisition compared to the other two conditions.

Collaboration Scripts to Support CSCL

To support individual knowledge acquisition in CSCL, group interactions, such as asking thoughtful questions (King, 2007), exchanging arguments (Kuhn, Shaw, & Felton, 1997), and engaging in exposing and reconciling cognitive discrepancies (Roschelle, 1992), have been shown to be effective. Yet, research has demonstrated that such processes do not generally occur spontaneously (e.g., Hewitt, 2005). Therefore, process-related scaffolding approaches have been developed to help learners perform more high-level, productive collaborative activities than without support. Collaboration scripts, as an example of such an instructional approach, have repeatedly been demonstrated to improve CSCL with respect to learners' knowledge acquisition, since they facilitate high-level collaboration processes by specifying, sequencing and distributing learning activities and roles among learners in a small group (Kollar, Fischer, & Hesse, 2006). For instance, Weinberger, Ertl, Fischer and Mandl (2005) used a social collaboration script, which distributed the roles of *case analyst* and *constructive critic* and sequenced sub-activities through *first case analysis*, *critique*, *reply to critique* to *final analysis* in an asynchronous CSCL environment and found that it fostered a variety of high-level online discussion processes and led to higher levels of domain-specific knowledge acquisition than unstructured CSCL.

Usually, collaboration scripts not only aim to foster the acquisition of domain-specific knowledge, but at the same time also domain-general skills (Wecker & Fischer, 2007). From an instructional point of view, focusing on learners' skill acquisition helps answer the question whether the effects of collaboration scripts extend beyond the learning situation in which they were provided, by promoting the capabilities of learners to collaborate in a fruitful way. By using collaboration scripts, the script itself may be internalized so that learners can become self-regulated in their use of it, since according to Vygotskian thinking, the actions of the roles and any verbal prompts can be internalized as inner speech so that learners can apply their acquired knowledge to self-prompt actions in similar situations (Rogoff, 1990). The empirical study presented below investigated the effects of script support on both learners' acquisition of domain-specific knowledge and their acquisition of domain-general skills; and more specifically the possible advantage of adaptable scripting in CSCL.

Adaptable Scripting to Foster Knowledge and Skill Acquisition

Although empirical evidence on the positive effects of classical CSCL scripts on individual knowledge and skill acquisition is ample, scripting approaches have also been criticized for being coercive and not allowing students to self-regulate their learning (Diziol, Walker, Rummel, & Koedinger, 2010). On the one hand, it often seems necessary to provide learners with scripts that impose some structure to enable them to engage in productive interaction. On the other hand, learners may be given too little freedom in scripted CSCL for deeper learning to occur (Rummel & Spada, 2005). Over the years, some efforts towards more open and less restrictive forms of CSCL scripts have been made to optimize the success of the scripting approach. In order to meet individual learners' needs for instructional support and their needs for freedom within a CSCL environment, adaptable instruction is a promising approach which may improve the fit between learner and learning environment. A learning environment is called 'adaptable' when users (i.e. teachers or students) can adjust external support to meet the learners' self- or other-perceived needs (Leutner, 2009). In the field of CSCL, there is a conceptual analysis of the notion of "flexibility", which deals with the concern of coercive scripting (Dillenbourg & Tchounikine, 2007). There, adaptable scripting was argued theoretically as a promising approach to realize flexibility. Whether the argument empirically holds true is one of the questions addressed in the study presented in this paper. Of course, there are reasons to expect this: by increasing learners' opportunities for self-regulation, adaptable scripting might be beneficial for learners' knowledge acquisition, since self-regulation has been regarded as an influential factor in learning, especially in skill acquisition (Zimmerman, 2008).

Research Questions and Hypotheses

This study aimed to answer the question to what extent an adaptable script can facilitate (1) the individual acquisition of *domain-specific knowledge* and (2) the individual acquisition of *domain-general skills* compared to a non-adaptable script and unscripted collaboration in a CSCL environment. We expected that learning by aid of an adaptable script would lead to higher levels of domain-specific knowledge and domain-general skills than a non-adaptable script, which in turn was expected to lead to better results than unstructured CSCL.

Besides the potential advantage of adaptable scripting on learning outcomes, we were also interested in its effects on learning processes: Therefore, we asked (3) how learners in the adaptable script condition adapted the script compared to the non-adaptable script condition; and (4) how the adaptation in turn affected their script adherence. We expected that learners in the adaptable script condition would do more self-planning and use less support provided by the script than learners in the non-adaptable script condition. The adaptation, yet, would not diminish their script adherence, compared to the non-adaptable script condition.

Method

Participants and Design

To answer these research questions, a one-factorial pre-post test design with three experimental conditions was used. The independent variable was the kind of instructional support. In the unscripted CSCL condition, students collaborated in an asynchronous CSCL environment without an external script. In the non-adaptable script condition, students were guided by a continuous, non-adaptable script, while students who learned in the adaptable script condition could adapt the script to their self-perceived needs (see below for more details). 54 students of Educational Sciences from the Ludwig-Maximilians-University of Munich participated in this study as part of a course assignment. They were randomly assigned to small groups of three, and each group was randomly assigned to one of the three experimental conditions (see Table 1).

Table 1: Design of the experimental study.

Unscripted collaboration	Non-adaptable script	Adaptable script
n = 18 students (6 triads)	n = 18 students (6 triads)	n = 18 students (6 triads)

Task, Setting and Learning Environment

The students' task was to analyze three authentic educational problem cases on the basis of Weiner's (1985) attribution theory. Discussions were led in a web-based CSCL environment, which was a revised version of the CASSIS environment (see Figure 1; Stegmann, et al., 2007). CASSIS is an asynchronous discussion board in which three participants can post messages that, apart from the experimenters, only the members of the learning group could read. The participants were logged in with code names in an effort to warrant anonymity.

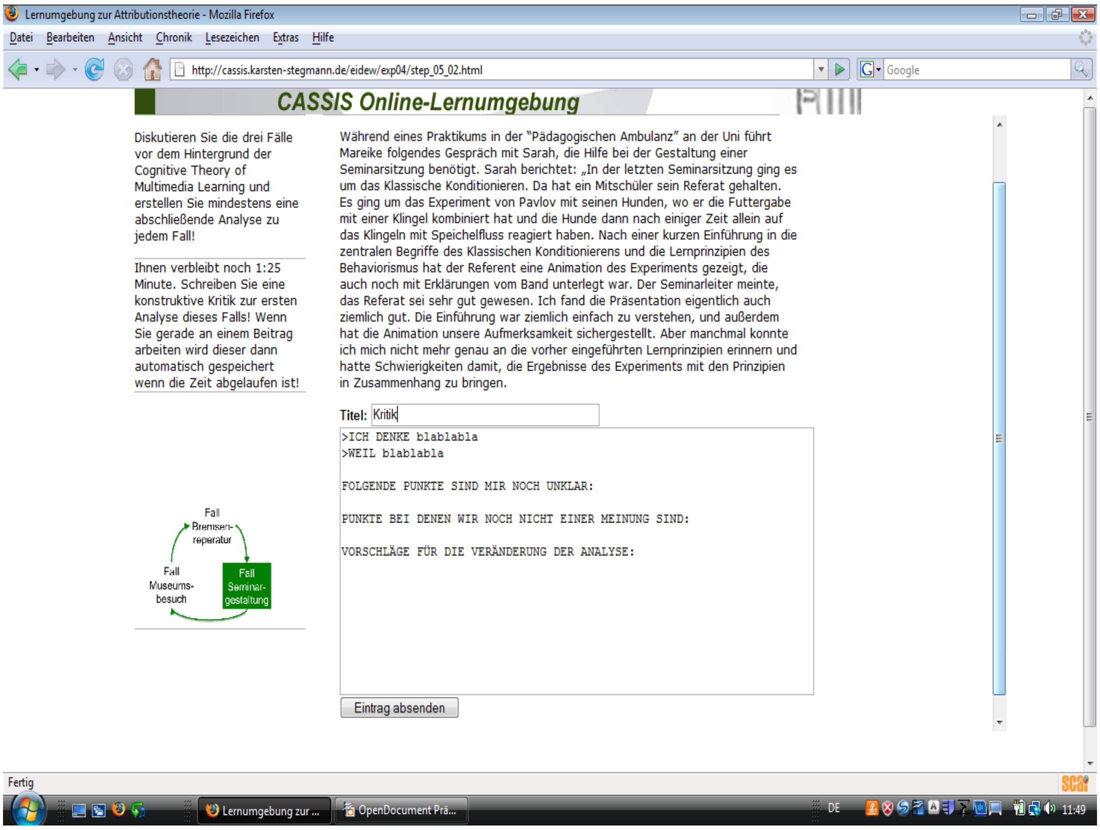


Figure 1. Screenshot of the learning environment CASSIS. Upper left: description of the task (analyze three problem cases with the help of the according theory). Middle left: timer (how much time left for the current task). Lower left: orientation map depicting which case the learner is working on. Upper right: case information. Lower right: discussion board where learners can type in messages (including script prompts).

Procedure and Experimental Conditions

The experiment spanned five phases. (1) *Pre-test and individual learning*: firstly, the participants of the study filled out questionnaires (demographic information and intrinsic motivation) and secondly they read two four-page theory sheets, one about Mayer's (2001) Cognitive Theory of Multimedia Learning and the other one about Weiner's (1985) Attribution Theory. The third step in the pre-test phase, which was the assessment of prior knowledge on collaboration (see below), took place immediately before the CSCL session, when the participants were arranged appropriately in the laboratory room.

(2) *Training*: after pre-test and individual learning, students were guided to a 44-minute training phase, which helped them get a first experience on how to handle the learning environment and how the collaboration script worked. During this phase, students in all three conditions analyzed three problem cases dealing with multimedia design against the background of Mayer's Cognitive Theory of Multimedia Learning, with the support of a collaboration script, which assigned two different roles (role A: analyst for one of the three cases and role B: constructive critic for the other two cases) to individual learners in a small group. Role A (analyst) took over the responsibility for the preliminary and concluding analysis on one case and responding to criticism from the learning partners. In the role of constructive critics (role B), learners were required to criticize their partners' analyses of the two other cases. These activities were supported by interaction-oriented prompts (e.g., "We have not reached consensus concerning these aspects:"), which were automatically inserted into the critics' messages and into the analyst's replies in order to help learners take over their roles successfully (Weinberger et al., 2005). In addition, there was a time limit for each sub-step students had to take.

(3) *Chat*: after a short break, students were guided to apply Weiner's attribution theory to three new problem cases. There was a chat before group work on analyzing the three cases, within which participants were asked to reflect upon their group discussion during training phase and plan for the coming one. In addition, students in the adaptable scripted condition were provided with opportunities to choose the role (analyst or critic) they would like to play for analyses of each problem case.

(4) *Treatment*: after 4 minutes of chatting, students were guided back to the same forum as they worked with during the training phase. In the 70-minute experimental phase, students in the unscripted condition

collaboratively worked on the case analyses without support of external scripts, but the strategies that had been discussed during chat could be used. Students in the non-adaptable script condition learned with the same script as they did during training. Here students could use the strategies that they discussed about only within the boundaries of the script. For students who learned in the adaptable script condition, the collaboration script in the experimental phase was adaptable. "Adaptability" was operationalized by providing learners with control over whether they would like to use the interaction-oriented prompts and the time they would like to spend on each case (time for the whole experimental phase kept the same as that in the other two conditions). Moreover, distribution of responsibilities for case analyses in the adaptable script condition was not determined by the script, but was based on their group decision during chat.

(5) *Post-test*: the experimental phase was followed by an individual post-test on students' knowledge acquisition on attribution theory, in which students were supposed to analyze a new problem case related to attribution theory, and a collaborative post-test on students' acquisition of collaboration skills, in which students were supposed to engage in an unstructured discussion about the post-test case (see below).

Dependent and Control Variables

Dependent variables were individual students' acquisition of domain-specific knowledge (on attribution theory) and individual students' acquisition of domain-general skills (collaboration skills). *Domain-specific knowledge acquisition* was assessed by an in-depth analysis of the individual analyses of a new problem case related to attribution theory during the post-test phase. Two coders independently first segmented 10% of individual analyses of the case into meaningful pieces of messages (Chi, 1997). Inter-rater agreement on segmentation was 94%, amounting to a Cohen's Kappa of κ = .88. The remaining 90 % of the material were then coded by the trained coders individually (the same procedure was applied to all the discourse analyses below). Secondly, each of the resulting segments was coded with the help of a coding system assessing the quality of knowledge construction (developed by Weinberger & Fischer, 2006). This system differentiated between utterances that represented the "construction of problem space" (e.g., "The student in the case thought that she failed in an exam because of low ability."), the "construction of conceptual space" (e.g., "Internal stable attribution of failure has negative effects on learning motivation.") and the "construction of relations between conceptual and problem space" (e.g., "The student is attributing internally stable when she took ability as the reason of her failure."). Inter-rater agreement on coding was 94%, amounting to a Cohen's Kappa of κ = .73. For the current analyses, only the frequency of segments coded as "construction of relations between conceptual and problem space" were included, since this category can be considered as representing the highest level of quality of knowledge construction.

Students' *acquisition of domain-general collaboration skills* was measured by a performance test, which was based on the unstructured group discussion about the post-test case. Individual contributions to the group discussion were first segmented into pieces of messages the same way as it was done for segmenting the individual analyses of the post-test case. Second, each of the resulting segments was assessed with respect to the application of collaboration skills that were introduced by the collaboration script. More specifically, the coding scheme differentiated whether a contribution represented "task specification" (e.g., "I suggest that we read and summarize our analyses in the discussion thread."), "role distribution" (e.g., "I am not sure with your point of view that external attribution is positive.") or "sequencing" (e.g., "We should first work on our own analyses."). The inter-rater agreement regarding coding on collaboration skills was 87%, and the inter-rater reliability was Cohen's κ = .71. The occurrence of these utterances was counted and the resulting sum scores were used as an indicator of individual acquisition of domain-general collaboration skills.

Beyond these outcome measures, we also used three process measures as dependent variables. To measure how learners in the adaptable script condition made use of the adaptability during their learning, their *self-planning* and *use of prompts* were coded. *Self-planning* was coded with the coding scheme from Zimmerman (2008). An utterance was coded as *self-planning* when the learners set goals and did strategic planning in terms of task coordination, time management etc. Since planning mainly takes place before the learning process, only learners' discussion during chat, which took place before the treatment phase, was coded in the current study. The inter-rater agreement regarding coding on learners' self-planning was 96 %, and the inter-rater reliability was Cohen's κ = .85. The indicator of learners' *use of prompts* during their online discussion was the percentage of prompts used by learners to prompts provided by the script. *Script adherence* as a third process variable was measured by counting if a specific sub-activity pre-described by the script showed up during the online discussion within a group. The script in the current study pre-described 8 sub-activities: first analysis, first critique from learning partner 1, first critique from partner 2, response to partner 1, response to partner 2, second critique from partner 1, second critique from partner 2, and final analysis.

As control variables, students' *prior knowledge on collaboration* was assessed individually during pre-test by an open format knowledge test. The test asked learners to describe how they would organize their group work imaging they were going to work in triads in an asynchronous discussion board – the CSCL scenario used in the current study; and more specifically which steps they would take and why. Students' responses were assessed with the same coding scheme that was used to assess the performance test on individual acquisition collaboration skills during post-test. The inter-rater agreement regarding coding on learners' prior knowledge on collaboration was 87 %, and the inter-rater reliability was Cohen's $\kappa = .71$. Furthermore, students' *initial motivation* was assessed after training by the motivation scale from Prenzel et al. (1993). This intrinsic motivation scale included five items (e.g., "During the learning session, I experienced myself as curious or inquisitive.") with Cronbach's $\alpha = .85$ in the reported study.

Results

Preliminary Analyses

Before performing statistical analyses related to our research questions, we checked whether the learners in the three experimental conditions were comparable with respect to prior knowledge and initial motivation. On none of these measures, we found significant differences in the pre-test (for prior knowledge on collaboration: $F_{(2,51)} = 0.11$; *n.s.;* for initial motivation: $F_{(2,51)} = 0.76$; *n.s.).* However, to avoid biases of effects of our treatment on the post test measures, we used these pre-test measures as control variables in all of the following analyses.

Individual Knowledge and Skill Acquisition

Descriptively, with respect to *domain-specific knowledge acquisition*, students who had learned with the adaptable script outperformed students from the other conditions (see table 2). An ANCOVA with the kind of instructional support as independent variable, group id as a random factor nested in treatment, the point scores on the individual knowledge test as dependent variable, prior knowledge on collaboration and initial motivation as control variables however indicated a non-significant effect for the kind of instructional support ($F_{(2,15)} = 3.30$; *n.s.*). The effect of group on individual acquisition of domain-specific knowledge was not significant ($F_{(15,35)} = 0.60$; *n.s.*). From Post-hoc-tests (LSD), it was found that students who had learned with the adaptable script slightly outperformed students from the non-adaptable script condition ($p = .08$). Students in the unscripted condition also performed slightly better than those from the non-adaptable scripted condition ($p = .11$).

Table 2: Descriptive values for individual knowledge and skill acquisition.

	Unscripted *M* (*SD*)	Non-adaptable script *M* (*SD*)	Adaptable script *M* (*SD*)	Total
Knowledge acquisition on attribution theory	7.86 (4.82)	5.58 (3.75)	8.12 (3.39)	7.19 (4.12)
Acquisition of collaboration skills	1.02 (1.49)	2.23 (1.59)	2.92 (2.16)	2.06 (1.91)

Paralleling the results on knowledge acquisition on attribution theory, with respect to individuals' acquisition of *domain-general collaboration skills*, the descriptive values depicted in Table 2 showed that students who had learned with the adaptable script outperformed students from the other two conditions. An ANCOVA with the kind of instructional support as independent variable, group id as a random factor nested in treatment, individual acquisition of collaboration skills as dependent variable, prior knowledge on collaboration and initial motivation as control variables indicated a significant effect of the kind of instructional support ($F_{(2,15)} = 3.75$; $p < .05$). The effect of group on individual skill acquisition was not significant ($F_{(15,34)} = 1.57$; *n.s.*). From Post-hoc-tests (LSD), it was found that students' acquisition of collaboration skills in the unscripted condition was substantially lower than in the adaptable script condition ($p < .01$) and slightly lower than in the non-adaptable condition ($p = .06$). There was no statistically significant difference between the non-adaptable and the adaptable script condition ($p = .16$).

Learners' Use of Adaptability

To answer the third research question about learners' use of adaptability in the adaptable script condition, compared to the non-adaptable script condition, descriptive values about learners' self-planning and their use of prompts were depicted in Table 3. Descriptively, data in table 3 showed that in the adaptable script condition, students did more self-planning and used less prompts than in the non-adaptable condition.

An ANOVA with the kind of instructional support (non-adaptable script vs. adaptable script) as independent variable, group id as a random factor nested in treatment, students' self-planning during chat (frequency) as dependent variable indicated a significant effect of the kind of instructional support ($F_{(1,10)} = 21.72$; $p < .01$). The effect of group on learners' self-planning was not significant ($F_{(10,24)} = 1.38$; *n.s.*). Since the *use of prompts* was calculated on group level, the procedure that AN(C)OVA with group id as a random factor nested in treatment was not applied. An ANOVA with the kind of instructional support (non-adaptable script vs. adaptable script) as independent variable, group use of prompts (percentage) as dependent variable indicated a significant effect of the kind of instructional support ($F_{(1,10)} = 4.95$; $p < .05$).

Table 3: Learners' use of adaptability.

	Non-adaptable script *M* (*SD*)	Adaptable script *M* (*SD*)
Self-planning	1.39 (1.33)	3.72 (1.36)
Use of prompts	0.31 (0.15)	0.14 (0.12)

Script Adherence

As students in the adaptable script condition were provided with opportunities to switch off the prompts, it ran the risk that they did not follow the scripts well. Analyses of script adherence answered the question how well students in each condition performed the eight sub-activities (first analysis, first critique from learning partner 1, first critique from partner 2, response to partner 1, response to partner 2, second critique from partner 1, second critique from partner 2, and final analysis) pre-described by the scripts. Figure 2 showed that groups in the adaptable script condition engaged overall less in the activities specified by the script than those in the non-adaptable script condition, especially with respect to sub-activity 2 (first critique from learning partner 1) and 3 (first critique from learning partner 2).

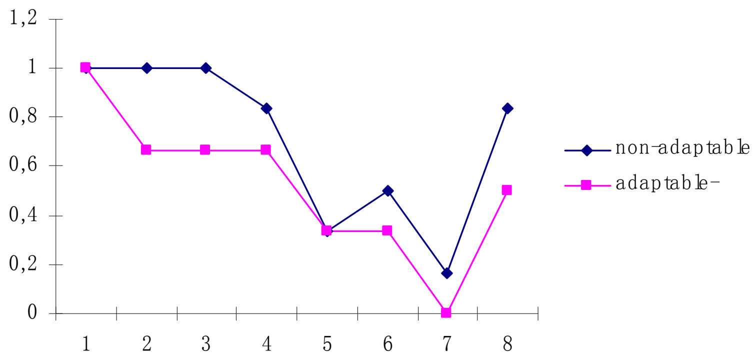


Figure 2. Script Adherence in Non-adaptable and Adaptable Script Condition.

An ANOVA with the kind of instructional support as independent variable and script adhenrence (percentage) as dependent variable indicated a significant effect of the kind of instructional support ($F_{(1,10)} = 7.33$; $p < .05$).

Discussion

In order to benefit from collaboration, students need to interact in productive ways. Collaboration scripts, as a process-related instructional approach, aim at inducing productive interaction in collaborative learning and therefore may improve learning outcomes. However, there are contradictory findings about the effects of collaboration scripts on individual learning outcomes. Many studies (e.g., Kollar et al., 2007) found positive results of CSCL scripts on individual knowledge acquisition, while a few others also (e.g., Weinberger et al., 2005) found negative effects. Against the background of related literature, negative script effects may be interpreted as caused by too high degree of coercion (Diziol et al., 2010). How to implement flexible scripts and thus to maximize the effectiveness of the scripting approach has recently drawn more and more attention. Therefore, we aimed to answer the question whether adaptable scripting is a promising approach to reduce the coercion of scripts without losing the benefit from the process-related support provided by scripts.

With respect to individual acquisition of domain-specific knowledge, the collaboration script used in the current study had slightly negative effects compared to the unscripted CSCL condition. However, when learners were provided with opportunities to reduce the structure imposed by the collaboration script by adapting some parts of the script according to their perceived needs, i.e. in the adaptable script condition, individual knowledge acquisition was enhanced when compared to the non-adaptable script condition. Regarding individual acquisition of domain-general collaboration skills, the collaboration script had slightly positive effect compared to the unscripted condition. When the script was adaptable, however, individual skill acquisition was substantially enhanced when compared to the unscripted condition. This supports the notion of Dillenbourg and Tchounikine (2007), who argued that "scripts must be flexible" (p.6). The adaptable script in the current study structured the collaboration by specifying and sequencing sub-activities, but left it modifiable to the students, for instance, to choose different roles and switch the prompts on and off. As shown in the results, students in the adaptable script condition benefited from their use of adaptability with respect to their acquisition of both domain-specific knowledge and domain-general skills. Adaptable scripting is therefore a promising approach to realize "flexibility" in scripted CSCL.

Beyond the "flexibility" perspective, positive effects of the adaptable scripting approach on individual learning outcomes may also be explained from the perspective of self-regulated learning (SRL). In line with related research, self-regulation is a significant predictor of students' academic outcomes, since self-regulated students would cognitively and motivationally engage in setting learning goals, using learning strategies and self-monitoring (Zimmerman, 2008). Further analyses of the learning process, which are currently under way, are required to confirm the extent to which the positive effects of adaptable scripting could be explained from the SRL point. For instance, we are interested to find out to what extent adaptability increases SRL, which has been linked to learning outcomes in CSCL (Dillenbourg, Järvelä, & Fischer, 2009; Hadwin, Oschige, Gress, & Winne, 2010).

Regarding the adaptation process, results showed that learners in the adaptable script condition did use the adaptability: they did more self-planning before online discussion and switched more prompts off during online discussion, compared to the non-adaptable script condition. As a consequence of adaptation, however, groups in the adaptable script condition engaged less in a script-like discussion process than groups in the non-adaptable script condition. Nevertheless, the overall decrease of script adherence caused by learners' adaptation of the script did not result in learners' losing the benefit from scripts. More data-sets and further analyses are required to distinguish effective adaptation groups from non-effective ones to examine the prerequisite for successful adaptation in scripted CSCL.

Limitations. The reported study has some limitations. Firstly, the sample size was relatively small. Statistics were done with 54 participants, 18 in each experimental condition. The relatively small sample size leads to constraints concerning statistical power. Therefore, the interpretation of the results should be made with caution. In order to enlarge the statistical power of this study, we are currently continuing data collection. Secondly, it is surprising that - although the same collaboration script was implemented both in the current study and a previous one (Weinberger et al., 2005), we found a slightly negative effect of the collaboration script on individual acquisition of domain-specific knowledge, whereas the previous study found a positive effect of this script on individual acquisition of domain-specific knowledge compared to unscripted CSCL. One main difference between the previous study and the current one is that in the previous study there was no training phase preceding the CSCL session, whereas in the current study, there was a 44-min training phase in advance. A possible explanation for this missing effect could be that the training phase led to some sort of internalization of the script so that learners in the unscripted condition still used parts of the script in the treatment phase. Another explanation from a motivational point of view would be the argument from Oehl and Pfister (2008), which states that scripts implemented in a relatively long term may elicit reactance responses. Possibly, further analyses on the learning processes provide evidence for or against these explanations.

Implications for future research and practice. Despite the limitations discussed above, the adaptable script in the current study is a practical example of realizing flexibility in scripted CSCL, but not the only way. Further ways how to implement flexibility in scripted CSCL may be differentially successful. For example, another line of research on flexible instruction is research on "adaptivity" (as opposed to adaptability realized in our study). An ongoing research (Mu, Stegmann, Mayfield, Rosé, & Fischer, 2010) which aims at realizing automatic adaptation through natural language processing (NLP) technique may provide evidence for the success of a different operationalization of flexibility in scripted CSCL. Flexible instruction, not only in CSCL, but also in other learning environments (e.g., classroom) is an issue that should be investigated in future studies. Finally, the current study showed that learners in the adaptable script condition benefited overall from the adaptability. Nevertheless, it is still an open question whether some students take more advantage out of adaptability than others do. Future studies may look for individual learning prerequisites that may moderate the positive effects of adaptability on knowledge acquisition, like, for instance, individual learning style.

References

Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *Journal of the Learning Sciences, 6(3)*, 271-315.

Dillenbourg, P. & Tchounikine, P. (2007). Flexibility in macro-scripts for computer-supported collaborative learning. *Journal of Computer Assisted Learning, 23,* 1-13.

Dillenbourg, P., Järvelä, S. & Fischer, F. (2009). The evolution of research on computer-supported collaborative learning: from design to orchestration. In N. Balacheff, S. Ludvigsen, T. de Jong, A. Lazonder, & S. Barnes (Eds.), *Technology-enhanced learning: principles and products*. New York: Springer.

Diziol, Walker, Rummel, & Koedinger (2010). Using intelligent tutor technology to implement adaptive support for student collaboration. *Educational Psychology Review, 22(1)*, 89-102.

Hadwin, A. F., Oschige, M., Gress, C. L. Z., & Winne, P. H. (2010). Innovative ways for using gStudy to orchestrate and research social aspects of self-regulated learning. *Computer in Human Behavior, 26,* 806-814.

Hewitt, J. (2005). Toward an understanding of how threads die in asynchronous computer conferences. *The Journal of the Learning Sciences, 14(4)*, 567-589.

King, A. (2007). Scripting collaborative learning processes: A cognitive perspective. In F. Fischer, I. Kollar, H. Mendl, & J. M. Haake. (Eds.) *Scripting computer-supported collaborative learning: Cognitive, computational and educational perspectives* (pp.13-37). New York, Springer.

Kollar, I., Fischer, F. & Hesse, F.W. (2006). Collaboration scripts - a conceptual analysis. *Educational Psychology Review, 18(2)*, 159-185.

Kollar, I., Fischer, F., & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning and Instruction, 17*,708-721.

Kuhn D., Shaw, V., & Felton, M. (1997). Effects of dyadic interaction on argumentative reasoning. *Cognition and Instruction, 15,* 287-315.

Leutner, D. (2009). Adaptivität und Adaptierbarkeit multimedialer Lehr- und Informationssysteme [Adaptivity and adaptability in multimedia learning and information systems]. In L. J. Issing, P. Klimsa (Eds). *Online-Lernen: Handbuch für Wissenschaft und Praxis*(pp.115-124). München: Oldenbourg.

Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.

Mu, J., Stegmann, K., Mayfield, E., Rosé, C., & Fischer, F. (2010, submitted). Framework for the development of classification schemes for automatic classifications of online discussions.

Oehl, M. & Pfister, H. R. (2008). Knowledge acquisition and problem solving in CSCL with learning protocols: Effects of different scripting strategies. In C. Bonk, M. M. Lee, & T. Reynolds (Eds.), *Proceedings of world conference on E-learning in corporate, government, healthcare, and higher education 2008* (pp. 3040-3045). Chesapeake, VA: AACE.

Prenzel, M., Eitel, F., Holzbach, R., Schoenhein, R.-J. & Schweiberer, L. (1993). Lernmotivation im studentischen Unterricht in der Chirurgie [Learning motivation in university-level surgery courses]. *Zeitschrift für Pädagogische Psychologie, 7*, 125-137.

Raudenbush, S.W., Bryk, A.S. (2006). *Hierarchical Linear Models: Applications and Data Analysis Methods.* Sage Publications, Inc.

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. New York: Oxford University Press.

Roschelle, J. (1992). Learning by collaboration: Convergent conceptual change. *The Journal of the Learning Sciences, 2,* 235-276.

Rummel, N., Spada, H. (2005). Learning to collaborate: an instructional approach to promoting collaborative problem solving in computer-mediated settings. *The Journal of the Learning Sciences, 14 (2)*, 201-224.

Stegmann, K., Kollar, I., Zottmann, J., Gijlers, H., de Jong, T., Dillenbourg, P., Slotta, J. D., & Fischer, F. (2007). Towards the convergence of CSCL and inquiry learning: Scripting collaborative inquiry learning. In C.Chinn, G. Erkens & S. Puntambekar (Eds.), *Mice, minds, and society. Proceedings of the CSCL 2007. Computer Supported Collaborative Learning* (pp. 829-830). Rutgers, The State University of New Jersey, New Brunswick, NJ, USA: International Society of the Learning Sciences.

Wecker, C., & Fischer, F. (2007). Fading scripts in computer-supported collaborative learning: The role of distributed monitoring. In C.Chinn, G. Erkens & S. Puntambekar (Eds.), *Mice, minds, and society. Proceedings of the CSCL 2007. Computer Supported Collaborative Learning* (pp. 764-772). Rutgers, The State University of New Jersey, New Brunswick, NJ, USA: International Society of the Learning Sciences.

Weinberger, A., Ertl, B., Fischer, F., & Mandl, H. (2005). Epistemic and social scripts in computer-supported collaborative learning. *Instructional Science, 33 (1)*, 1-30.

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education, 46,* 71-95.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review, 92*, 548-573.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developents, and future prospects. *American Educational Research Journal, 45(1)*, 166-183.

Explaining the Effects of Continuous and Faded Scripts on Online Search Skills: The Role of Collaborative Strategy Practice

Christof Wecker, Ingo Kollar, Frank Fischer, Ludwig-Maximilians-Universität München, Leopoldstr. 13, 80802 München, Germany
Email: christof.wecker@psy.lmu.de, ingo.kollar@psy.lmu.de, frank.fischer@psy.lmu.de

Abstract: It has been shown that continuous and faded collaboration scripts can foster important components of scientific literacy such as online search skills. The present paper studies the effects of these types of scripts on learners' practice of the strategy suggested by the script, as well as the relation of the learners' and their learning partners' practice of the strategy to the development of online search skills. Data from a four-week field study with 129 ninth-grade students in three conditions (no script, continuous script and faded script) were analyzed. Findings indicate positive effects of both scripts compared to unsupported collaboration on the practice of the strategy. Learners' own, but not their learning partners' practice of the strategy predicted their development of online search skills. These results indicate that neither mere exposure to the script nor observation of learning partners performing the strategy explain the effects of scripts on learning outcomes.

Prior Research on Collaboration Scripts

In recent years, collaboration scripts have become a main focus of research on computer-supported collaborative learning. Collaboration scripts are socio-cognitive scaffolds that specify activities, sequence them and distribute them among different roles taken over by members of a small group of learners, with the aim to help students advance their domain-specific knowledge about the topic of collaboration and domain-general skills such as the ones that also belong to scientific literacy (Kollar, Fischer & Hesse, 2006). Several studies have demonstrated positive effects of scripts on learning (e.g., De Wever, Van Keer, Schellens, & Valcke, 2009; Kollar, Fischer & Slotta, 2007; Rummel, & Spada, 2005; Slof, Erkens, Kirschner, Jaspers & Janssen, 2010; Stegmann, Weinberger & Fischer, 2007; Weinberger, Stegmann & Fischer, 2010).

Some studies of support for collaborating learners other than collaboration scripts have been conducted in field settings (e. g. Hmelo-Silver, 2006; Walker, Rummel & Koedinger, 2009). However, most research on effects of computer-supported small-group collaboration scripts on learning *outcomes* is done in the laboratory so far. Some lab studies demonstrated beneficial effects of different kinds of collaboration scripts on domain-*general* skills such as argumentation skills during online discussions in a problem-based learning context (Stegmann et al., 2007). In a further study in the context of web-based inquiry learning, a collaboration script that was designed to help students formulate well-grounded arguments, counterarguments and integrations and implemented at specific points of a WISE curriculum unit helped students develop higher levels of the domain-general skill of argumentation, but did not help them reach higher levels of domain-specific knowledge when compared with less structured collaboration (Kollar et al., 2007). Moreover, it appeared that during the learning process the collaboration script only raised argumentation quality as long as it was present. As soon as the script was "switched off", learners relapsed into their previous argumentation style. In other words, development of robust argumentation skill hardly occurred, which may be explained by a rather short learning time of 120 minutes.

To investigate whether computer-supported collaboration scripts produce more robust effects when longer learning phases are studied, we conducted a larger field study in an inquiry curriculum unit in Biology (Wecker, Kollar, Fischer & Prechtl, 2010). One of the main purposes of Science Education is to prepare all students for their role as responsible citizens including participation in societal debates that involve scientific issues, which requires both fundamental knowledge about important scientific principles and the ability to gather and evaluate more specific and recent information that goes beyond what can ever be learned in school. Therefore we selected domain-specific knowledge from Biology and online search skills as the two targeted learning outcomes of our intervention. The curriculum unit was implemented in regular class periods led by the classes' Biology teachers. It comprised seven lessons about Genetic Engineering and lasted about four weeks. A collaboration script for collaborative online search was compared to unscripted collaboration in an experimental design and produced substantial positive effects with respect to the two learning outcomes (Wecker, Kollar et al., 2010). In particular with respect to online search skills, this effect could in principle be due to mere exposure to the prompts included in the collaboration script during the test session rather than their application during collaborative learning. In this paper we present further analyses based on data from this study to clarify the mechanisms behind the beneficial effect of a collaboration script on online search skills. It was assumed that it is the actual practice of the strategy suggested by the collaboration script rather than mere exposure to the prompts that fulfils this role or observation of a learning partner performing the strategy. If the effects of collaboration scripts on online search skills are mainly due to mere exposure to the prompts contained in the script, the amount of practice of the strategy by the learners themselves should be unrelated to their online search skills in the posttest. If learners develop online search skills mainly because they observe their learning partners perform the strategy suggested by the script, the amount of the learners' partners' practice of the strategy should be related to their own online search skills in the posttest. These potential explanations for script effects are tested by means of the analyses presented in this paper.

A further goal of the study was to investigate the effect of a faded collaboration script compared to the more "classical" continuous one. Fading is the gradual removal of support as learners become more proficient and can take over control of increasing parts of the activity to be learned. Thereby learners get the opportunity to practice not only the details of the execution, but also their regulation. Hence, fading is considered as part and parcel of scaffolding (Pea, 2004; Puntambekar & Hübscher, 2005; Wood, Bruner & Ross, 1976). So far, research on the effects of fading of scaffolds in general is sparse, and the results are mixed. In one of two experiments, Leutner (2000) found evidence for beneficial effects of fading, whereas in the other one performance decreased in the process of fading. In another study a marginally significant positive effect of fading was found with respect to only one of several aspects of knowledge about the principles of scientific explanations (McNeill, Lizotte, Krajicik & Marx, 2006).

Transferring the concept of "fading" to research on collaboration scripts, Wecker and Fischer (2007; 2010) could show in an empirical study in a computer-supported collaborative learning context that the fading of an instructional script alone had no positive effect on learning. If it was combined with "distributed monitoring" (cf. King, 1998), i. e. when the learning partner was asked to monitor whether his or her peer complied with the strategy induced by the script, however, a significant positive effect on students' strategy knowledge in comparison to a continuous script was found. Just like the study by Kollar et al. (2007), however, this study was conducted as a lab experiment with a rather short learning time.

In our field experiment in Biology lessons, we investigated whether a faded collaboration script had a positive effect on the acquisition of online search skills when implemented over a more extended time frame in a field setting. The script contained meta-cognitive aspects such as monitoring of the learning partner's performance as part of the script itself that were faded along with the whole script, accordingly. A positive effect of the faded collaboration script compared to unscripted collaboration was in fact found with respect to domain-general online search skills, but not with respect to domain-specific knowledge. However, the fading of the collaboration brought no additional effect beyond that of a continuous collaboration script with respect to online search skills (Wecker, Kollar et al., 2010). In the present paper we investigate whether this finding can be explained by the learners' practice of the strategy during the fading of the collaboration script.

Research Questions

(1) How are students' and their learning partners' *practice of the strategy* suggested by the script in the context of collaborative inquiry learning in the classroom related to their acquisition of *domain-specific knowledge* and *online search skills*?

We hypothesized that the learners' own practice of the strategy during collaborative learning is more strongly related to the acquisition of online search skills than their learning' partners practice of the strategy.

(2) What are the effects of continuous and faded small-group collaboration scripts for collaborative online search on students' *practice of the strategy* in the context of collaborative inquiry learning in the classroom?

We expected that both a continuous and a faded collaboration script have a positive effect on the practice of the strategy. We did not expect a difference between these two types of scripts in this respect.

Method

Participants and Design

Participants in this study were 129 students from six ninth-grade classes from three urban high schools. Their average age was 14.7 years (SD = 0.67); 60 of them were female, 69 were male. Data from the students were only collected if their parents had agreed individually that their child was included in the data collection.

A one-factorial quasi-experimental design with three experimental conditions differing in the type of instructional support for collaborative online search was implemented with the classes as the units of random assignment (see table 1). One condition without collaboration script, one with a continuous collaboration script, and one with a faded collaboration script were compared.

Table 1: Design of the study.

Type of instructional support		
No collaboration script	*Continuous collaboration script*	*Faded collaboration script*
N = 36 students from 2 classes	N = 46 students from 2 classes	N = 47 students from 2 classes

Instructional Setting and Procedure

The study was conducted in an inquiry-based curriculum unit that spanned seven regular biology lessons, which were preceded and followed by a test session, respectively. The students' task was to arrive at a decision about whether they supported or rejected the genetic modification of plants for the purpose of food production. The lessons were led by the classes' regular biology teachers.

After a short introduction to relevant background domain knowledge from Genetics, the students in all three conditions received the same introduction to the strategy of online search. Lessons two through seven consisted of three consecutive learning cycles about three different topical aspects of the discussion about Genetic engineering of plants: one on economic, one on ecological, and one on health aspects. Each cycle comprised three consecutive steps. First, students in dyads collaboratively browsed through an online library in order to gather background information about Genetics and Genetic Engineering relevant to the current aspect of the discussion. In the second step, student dyads conducted collaborative online searches to elaborate or modify their initial argument concerning the current aspect of the discussion. In these phases of collaborative online search, the independent variable was manipulated (see below). In the third step, students exchanged their arguments in a plenary discussion based on the findings from their collaborative online searches.

The online library was implemented as a module in the Web-Based Inquiry Environment (WISE; Slotta & Linn, 2009). The students in each dyad sat next to each other in front of their two computers and communicated face-to-face, but each student had a laptop computer with a mouse for him- or herself. During the collaborative online search phases, the browsers of the students in each dyad were connected via a software tool called S-COL (Wecker, Stegmann et al., 2010). This allowed for collaborative Internet browsing, i.e., during their online searches, both learning partners from each dyad always saw the same web pages, no matter who of them clicked on a link or used the navigation functions of their browser.

Independent Variables

No Collaboration Script

The students in the condition without a collaboration script received no support for their collaborative activities beyond an introduction by their teacher to the strategy of online search at the beginning of the curriculum unit. This information was equivalent to the information presented in the prompts of the collaboration scripts in the other two conditions.

Continuous Collaboration Script

This script was targeted at improving collaborative online search. That is, it was directed at the specific activities required to perform a successful online search, and it divided these activities among two learners in order to induce interaction concerning crucial aspects of the skill to be acquired (Dillenbourg & Jermann, 2007). It was implemented as follows: In addition to connecting the computers of the two partners of a dyad, the S-COL software tool described above was used in the two experimental conditions to display particular prompts on the basis of the type of website the students were accessing at any time (Google start page, Google hit list, any other web site). In each dyad there were two roles (A and B) that switched after returning to Google, which they were required to use for their searches, from any other web page encountered during the search activities. The collaboration script was implemented as complementary text prompts in the scaffolding areas of S-COL in the browsers of both group members (left part of the screen in figure 1). The script contained support for five – partly iterative – steps of collaborative online search: (1) the *formulation of an initial argument and a sketch of the information needed*, (2) the *selection of search terms*, (3) the *evaluation of the hit list*, (4) the *localization of relevant information* on a web page, and (5) a *written formulation of the final elaborated argument*. For example, during the second stage, the *selection of search terms*, learner A was prompted to perform the activities to suggest a set of search terms and discuss them with B, while B had the task to perform the activities to first recall the information they had decided to look for, and comment on A's suggestions for the search terms with respect to their likelihood of yielding suitable as well as inappropriate hits.

Faded Collaboration Script

In the condition with the faded collaboration script the same prompts as in the condition with the continuous collaboration script were used to support collaborative online search, but they were continuously removed, the more online searches the dyad performed. The general fading scheme was that after a series of four external web sites had been accessed, the prompts described above became more unspecific: Initially, the scaffolding area of S-COL contained both the names of the individual activities as well as explanatory text. At the second fading stage, only the names of the activities were displayed. At the final stage, only headings for the actual steps were displayed. In each of the three cycles of the curricular unit, the degree of support initially reverted to a slightly higher level before it was further reduced according to the scheme just described.

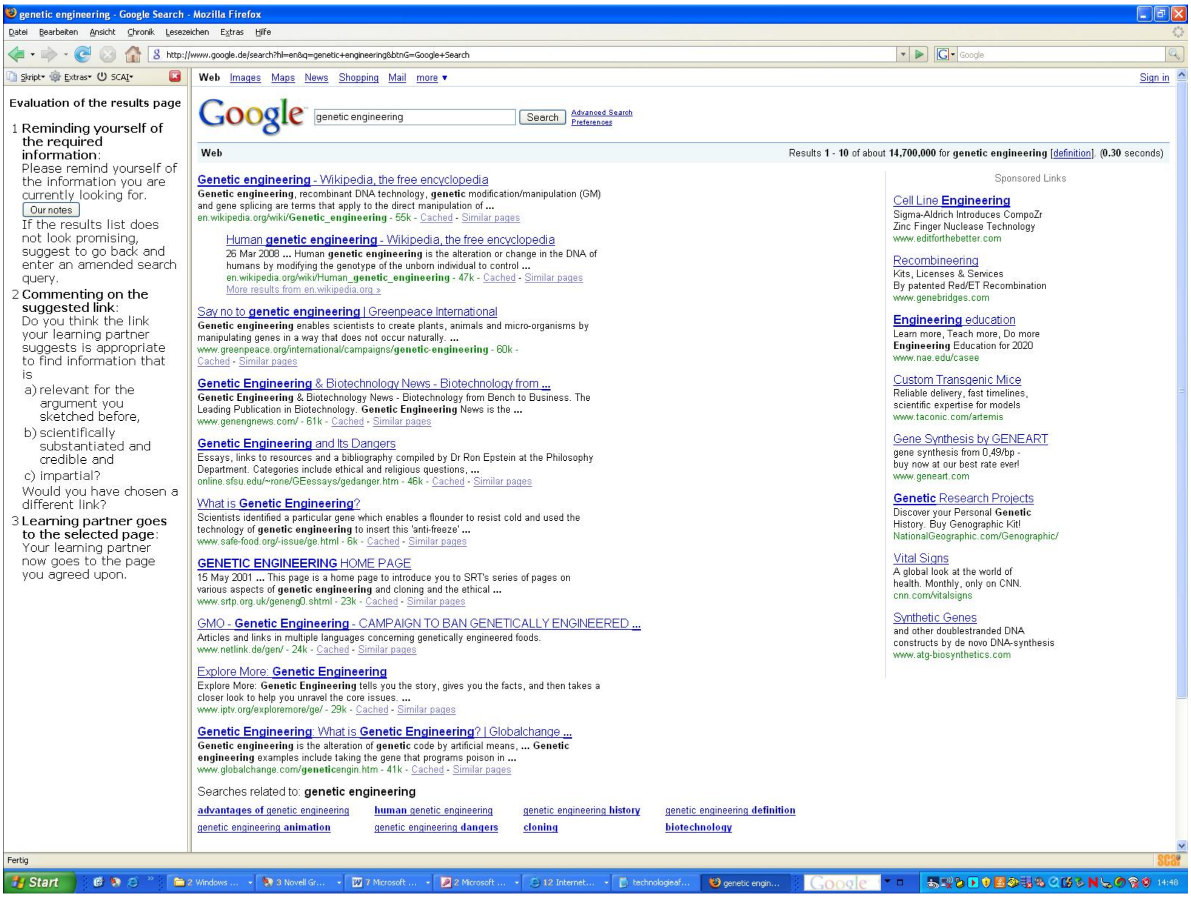


Figure 1. Part of the Continuous Collaboration Script Presented along with a Google Hit List.

Data Sources and Dependent Variables

Practice of the Strategy

Screen-and-audio capturing software was used to record both the learners' utterances during face-to-face collaboration in front of their computers and their activities on their computers. To deal with the vast amount of data, a time sample of 10 minutes from the beginning of the second step of each of the three inquiry cycles, i. e. the online-search, was selected for in-depth analysis from each dyad. The whole online-search phases from which these three 10 minutes time samples were drawn lasted for 45, 30 and 30 minutes, respectively.

To arrive at an indicator of the *practice of the strategy* suggested by the script, the occurrence of the activities suggested by the script was coded separately for both members of each dyad for segments of 10 seconds of length on the basis of the learners' as well as their partners' activities on their computers. A coding scheme with separate, mutually exclusive codes for all activities suggested by the collaboration script was applied to the segments of both learners in a dyad. That is, for each time segment, the coders identified for each of the two learners which activities he or she predominantly performed during this segment.

Because all time samples were drawn from the same stage of the online-search phases in each cycle, namely the beginning of the online-search phases, a subset of activities suggested by the script can be identified that are most appropriate in this early phase. These are the activities subsumed under the step "*formulation of an* *initial argument and sketch of the information needed*" (cf. the description of the stages in the section on the continuous collaboration script above). Therefore, the proportion of time spent on the activities belonging to this step was used as an indicator for the extent to which the learners performed the strategy.

The screen-audio-recordings were analyzed by three independent coders who were evenly distributed across the three conditions. A subsample of 11% of the data was coded by all three coders to determine their agreement, which was acceptable (Coders 1 and 2: 70% agreement; Cohen's $\kappa = .67$; Coders 1 and 3: 68% agreement; Cohen's $\kappa = .64$; Coders 2 and 3: 91% agreement; Cohen's $\kappa = .89$).

Based on these codings, the practice of the strategy was operationalized as the frequency of segments within the first ten minutes of each online search phase, in which one of the activities subsumed under the first step of the script was performed, relative to the total length of the time interval included in the analysis. The reliability of this variable as calculated over single variables indicating the relative frequency of the single activities subsumed under the first step of the script was good (Cronbach's $\alpha = .82$).

Online Search Skills

In the tests measuring online search skills, students were asked to describe in as much detail as possible how they would use the Internet to form a position about a specific sample issue without actually doing so. The answers were pre-structured by means of a two-column table with eight rows. The students were instructed to describe the single steps involved in the search in the left hand-column and the evaluative criteria to be applied in combination with these specific steps in the right-hand column. Pre- and posttest were coded for the occurrence of each individual element of a general ideal solution that contained the steps that were also triggered in the collaboration script as well as the evaluative criteria suggested in the script. Each student's responses were analyzed by one of two coders who were blind to condition. The proportions coded by each of them were equal across the three experimental conditions. An overall scale for *online search skills* was formed by counting the coded occurrences of steps and evaluative criteria. Separate scales for *knowledge about single steps* and *knowledge about evaluation criteria* were also formed.

For the assessment of inter-coder agreement, the two coders analyzed 15 % of the data from the pretest and 25 % of the data from the posttest independently of each other. For all participants included in this sample the values of the overall scale as well as the two subscales were calculated from the codings of both coders. Intra-class correlations for single measures amounted to *ICC* = .83 for the overall scale of *online search skills* in the posttest (*ICC* = .51 in the pretest), *ICC* = .82 for *knowledge about single steps* and *ICC* = .67 for *knowledge about evaluative criteria*.

Statistical Analysis

The significance level was set to 5 % for all analyses.

Results

Research Question 1: Relations between Students' and Their Learning Partners' Practice of the Strategy Suggested by the Script and Their Acquisition of Online Search Skills

The learners' practice of the strategy suggested by the script was significantly related to the overall scale of online search skills from the posttest, $r(N = 129) = .27$; $p < .01$ (one-sided), the subscale for knowledge about single steps required to perform a successful online search, $r(N = 129) = .20$; $p = .01$ (one-sided), and the subscale for knowledge about evaluation criteria, $r(N = 129) = .30$; $p < .01$ (one-sided). The same holds for *their learning partners'* practice of the strategy suggested by the script and its relation to the learners' score on the overall scale of online search skills from the posttest, $r(N = 129) = .26$; $p < .01$ (one-sided), the subscale for knowledge about single steps required to perform a successful online search, $r(N = 129) = .21$; $p < .01$ (one-sided), and the subscale for knowledge about evaluation criteria, $r(N = 129) = .26$; $p < .01$ (one-sided). Three regression analyses for the overall scale for online search skills and its two subscales in the posttest as criterion variables with learners' and their learning partners' practice of the strategy suggested by the script as predictors were conducted, while also controlling for prior online search skills. They showed that the learners' own practice of the strategy suggested by the script significantly predicts two of the learning outcomes (online search skills: $\beta = .25$; $p = .02$; knowledge about single steps: $\beta = .19$; $p = .09$; knowledge about evaluation criteria: $\beta = .27$; $p = .02$), while their learning partners' practice of the strategy suggested by the script does not (online search skills: $\beta = .01$; $p = .90$; knowledge about single steps: $\beta = .00$; $p = .99$; knowledge about evaluation criteria: $\beta = .04$; $p = .76$).

Research Question 2: Effects of Continuous and Faded Small-Group Collaboration Scripts on Students' Practice of the Strategy

While students in the condition without a collaboration script engaged in activities subsumed under the first step of the script 3.0 % (*SD* = 0.04) of the time during the first ten minutes of each online search phase, students in the condition with a continuous script performed these activities in 7.3 % (*SD* = 0.06) of the time and students in the condition with a faded script in 7.2 % (*SD* = 0.08). That is, the average time spent on activities related to the formulation of an initial argument and the sketch of the information needed within the first ten minutes of online search was doubled to more then 40 seconds by the use of one of the two collaboration scripts. An analysis of variance with the students' practice of the strategy suggested by the script as the dependent variable and the kind of instructional support and classes nested within the three kinds of instructional support as independent variables revealed a significant medium-size main effect of the type of instructional support, $F(2; 122) = 6.89$; $p < .01$; partial $\eta^2 = .10$.

Planned comparisons revealed significant effects of large size between the condition with continuous collaboration script and the condition with no collaboration script, $F(1; 77) = 15.44$; $p < .001$; partial $\eta^2 = .17$, and of medium to large size between the condition with faded collaboration script and the condition with no collaboration script, $F(1; 78) = 10.41$; $p < .01$; partial $\eta^2 = .12$. However, no significant difference was found between the condition with the faded collaboration script and the condition with the continuous collaboration script, F(1; 88) < 1; *n. s*.

Discussion

This field experiment provides evidence that collaboration scripts can foster learning activities that lead to the development of online search skills, which have to be regarded as an important component of scientific literacy. The results indicate in particular that the collaboration script approach (Kollar et al., 2006) could successfully be transferred to structure web-based collaborative inquiry learning in real secondary school classrooms. Typically in laboratory studies the degrees of adherence to the prompts of collaboration scripts (i. e. the proportion of prompts that are responded to appropriately) are as high as about 70 % (Stegmann et al., 2007, p. 434) or 90 % (Wecker & Fischer, 2010). We currently do not have comparable numbers for the present study, but the share of the learners' activities that corresponded to parts of the strategy suggested by the script was below 10 %. This finding particularly highlights the importance of structuring collaboration in educational fields of practice.

It could also be shown that the practice of the strategy suggested by the script during collaborative learning is related to the development of online search skills. It is conceivable that effects of collaboration scripts on online search skills are mainly due to mere exposure to the prompts contained in the script. If this were the case, the amount of practice of the strategy by the learners themselves should be rather unrelated to their online search skills in the posttest. Furthermore it is possible that learners acquire online search skills also because they observe their learning partners perform the strategy suggested by the script. If this explanation were correct, the amount of the learners' partners' practice of the strategy should be related to their own online search skills in the posttest. What we do find, however, is – contrary to the first assumption – that the learners' own practice of the strategy is correlated with their online search skills in the posttest, and that – contrary to the second assumption – the learners' own practice of the strategy significantly predicts online search skills in a multiple regression analysis with the learners' and their partners' practice of the strategy as predictor variables, whereas the learning partners' practice of the strategy does not. This constitutes evidence that in fact the learners' own practice might play the central role for the development of online search skills, and not just the exposure to the prompts of the script or the observation of a partner performing the strategy. It does not, however, constitute definitive evidence that the learning partner does *not* play a role as a model because the insignificant regression coefficient for the learning partners' practice of the strategy could also be due to multicollinearity. This issue should be picked up in future research.

Taken together, the pattern of results presented in this paper offers an explanation why a continuous and a faded collaboration script both have a positive effect on the development of online search skills, while there is no difference between the two script types in this respect. The development of online search skills is correlated with the learners' practice of the strategy suggested by the script, and the pattern of the scripts' effects on the practice of the strategy precisely mirrors the effects on online search skills (Wecker et al., 2010). This suggests the explanation that the two types of scripts affect the development of online search skills via their effects on the practice of the strategy contained in the script.

With respect to the effectiveness of the fading of the collaboration script, however, some aspects require further discussion. The domain-general skills of online search can apparently be positively affected by a faded collaboration script compared to unscripted collaboration. However, no additional effect of fading compared to a continuous collaboration script could be demonstrated. Although some aspects of distributed monitoring (Wecker & Fischer, 2007) were integrated in the collaboration script used in this study (e. g. asking learner B to keep an eye on the sketch of the information the group is currently trying to find), this may be not sufficient under field conditions, in particular if prompts for these metacognitive aspects are faded along with the other parts of the script. A functional equivalent for distributed monitoring that should be explored in future research may be feedback by a teacher pointing towards the dependency of success and failure in the collaborative task on the adherence to the collaboration script (cf. Schunk & Rice, 1993). Another difference between the present study and the laboratory study by Wecker and Fischer (2007) was the fact that the collaboration script embodied a collaborative strategy (of collaborative online search) that learners were expected to acquire as an individual strategy (of individual online search). Therefore, successful development of the corresponding domain-general skill on an individual level requires the integration of components of the strategy that are distributed over learners in a collaborative situation, which could be difficult without phases of individual practice (cf. Anderson & Lebiere, 1998). Accordingly, the successful development of a domain-general skill previously practiced collaboratively may require phases of *individualization* of the corresponding strategy. This can be regarded as another dimension of fading: Besides withdrawing technologically implemented prompts of a collaboration script, fading may also be constituted by a gradual shift from higher social levels (e.g., dyads) to the individual, which may be closer to the original ideas about scaffolding and fading (Wood, Bruner, & Ross, 1976; Pea, 2004). An example of such an approach that contains fading across social levels is reciprocal teaching (Palincsar & Brown, 1984), but to our knowledge, this aspect of fading has not been systematically varied in an experimental study so far.

In sum, it is still an issue for future research on collaboration scripts to explore under which conditions fading can help learners internalize a collaboration script and yield a positive effect in addition to the effect of collaboration scripts themselves. Candidates for these conditions could be approaches to embedding small-group collaborative learning in the overall instructional setting on different social levels (Dillenbourg & Jermann, 2007; Tabak, 2004), i. e. classroom scripts. Another study from the same project has shown that a classroom script involving phases of small-group collaboration combined with modelling on the plenary level can foster online search skills without the need to permanently structure small-group collaboration (Kollar, Wecker, Langer & Fischer, 2011). It is an open question how such a classroom script interacts with the fading of a small-group collaboration script.

The present paper extends our understanding of collaboration scripts by providing evidence that computer-supported collaboration scripts can be effective also in real-world settings and in interventions with an extended timeframe, and that similar mechanisms as the ones identified in laboratory studies can explain this (Kollar et al., 2007; Stegmann et al., 2007). Computer-supported collaboration scripts have been shown to be effective means to foster important components of scientific literacy such as online search skills. Apparently by means of computer-based collaboration scripts it is possible to create a zone of proximal development involving the learning partner as well as technology and thereby help learners build up domain-general skills supported by the external collaboration script (Carmien, Fischer, Fischer & Kollar, 2007; Pea, 2004; Wood et al.; 1976).

References

Anderson, J. R. & Lebiere, C. (1998). *The atomic components of thought*. Mahwah: Erlbaum.

Carmien, S., Fischer, F., Fischer, G. & Kollar, I. (2007). The interplay of internal and external scripts - a distributed cognition perspective. In F. Fischer, H. Mandl, J. M. Haake & I. Kollar (Hrsg.), *Scripting computer supported communication of knowledge: Cognitive, computational and educational perspectives* (S. 303-326). New York: Springer.

De Wever, B., Van Keer, H., Schellens, T., & Valcke, M. (2009). Structuring asynchronous discussion groups: The impact of role assignment and self-assessment on students' levels of knowledge construction through social negotiation. *Journal of Computer Assisted Learning, 25*, 177–188.

Dillenbourg, P. & Jerman, P. (2007). Designing integrative scripts. In F. Fischer, H. Mandl, J. M. Haake & I. Kollar (Eds.), *Scripting computer supported communication of knowledge: Cognitive, computational and educational perspectives* (pp. 275-301). New York: Springer.

Hmelo-Silver, C. E. (2006). Design principles for scaffolding technology-based inquiry. In A. M. O'Donnell, C. E. Hmelo-Silver & G. Erkens (Eds.), *Collaborative learning, reasoning, and technology* (pp. 147–170). Mahwah: Erlbaum.

King, A. (1998). Transactive peer tutoring: Distributing cognition and metacognition. *Educational Psychology Review, 10*, 57-74.

Kollar, I., Fischer, F., & Hesse, F. W. (2006). Computer-supported collaboration scripts - a conceptual analysis. *Educational Psychology Review, 18*(2), 159-185.

Kollar, I., Fischer, F. & Slotta, J.D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning and Instruction, 17*, 708-721.

Kollar, I., Wecker, C., Langer, S. & Fischer, F. (2011, July). *Orchestrating Web-Based Collaborative Inquiry Learning with Small Group and Classroom Scripts*. Paper presented at the CSCL 2011 "Connecting computer-supported collaborative learning to policy and practice", Hong Kong, 4th. – 8th July 2011.

Leutner, D. (2000). Double-fading support - a training approach to complex software systems. *Journal of*

Computer Assisted Learning, 16, 347-357.
McNeill, K. L., Lizotte, D. J., Krajcik, J. & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *Journal of the Learning Sciences, 15(2)*, 153-191.
Palincsar, A. S. & Brown, A. L. (1984). Reciprocal Teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1*(2), 117-175.
Pea, R. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *Journal of the Learning Sciences, 13(3)*, 423-451.
Puntambekar, S., & Hübscher, R. (2005). Tools for scaffolding students in a complex environment: What have we gained and what have we missed? *Educational Psychologist, 40 (1)*, 1-12.
Rummel, N., & Spada, H. (2005). Learning to collaborate: An instructional approach to promoting collaborative problem solving in computer-mediated settings. *The Journal of the Learning Sciences, 14*(2), 201–241.
Schunk, D. H. & Rice, J. M. (1993). Strategy fading and progress feedback: Effects on self-effeicacy and comprehension among students receiving remedial reading services. *The Journal of Special Education, 27*(3), 257-276.
Slof, B., Erkens, G., Kirschner, P. A., Jaspers, J. G. M., & Janssen, J. (2010). Guiding students' online complex learning-task behavior through representational scripting. *Computers in Human Behavior, 26*(5), 927–939.
Slotta, J. D. & Linn, M. C. (2009). *WISE Science: Web-based inquiry in the classroom*. New York, NY: Teachers College Press.
Stegmann, K., Weinberger, A. & Fischer, F. (2007). Facilitating argumentative knowledge construction with computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning, 2(4)*, 421-447.
Tabak, I. (2004). Synergy: A complement to emerging patterns of distributed scaffolding. *The Journal of the Learning Sciences, 13*(3), 305-335.
Walker, E., Rummel, N. & Koedinger, K. (2009). Integrating collaboration and intelligent tutoring data in evaluation of a reciprocal peer tutoring environment. *Research and Practice in Technology Enhanced Learning, 4*(3), 221-251.
Wecker, C. & Fischer, F. (2007). Fading scripts in computer-supported collaborative learning: The role of distributed monitoring. In C. Chinn, G. Erkens & S. Puntambekar (Eds.), *Proceedings of the CSCL 2007. Rutgers, The State University of New Jersey, New Brunswick, New Jersey, USA, July 16th – July 21st, 2007* (pp. 763-771). International Society of the Learning Sciences.
Wecker, C., & Fischer, F. (2010). Fading instructional scripts: preventing relapses into novice strategies by distributed monitoring. In K. Gomez, L. Lyons, & J. Radinsky (Eds.) *Learning in the Disciplines: Proceedings of the 9th International Conference of the Learning Sciences (ICLS 2010) - Volume 1, Full Papers*. International Society of the Learning Sciences: Chicago IL.
Wecker, C., Kollar, I., Fischer, F. & Prechtl, H. (2010). Fostering online search competence and domain-specific knowledge in inquiry classrooms: effects of continuous and fading collaboration scripts. In K. Gomez, L. Lyons, & J. Radinsky (Eds.) *Learning in the Disciplines: Proceedings of the 9th International Conference of the Learning Sciences (ICLS 2010) - Volume 1, Full Papers* (pp. 810-817). International Society of the Learning Sciences: Chicago IL.
Wecker, C., Stegmann, K., Bernstein, F., Huber, M. J., Kalus, G., Rathmeyer, S., Kollar, I. & Fischer, F. (2010). S-COL: A Copernican turn for the development of flexibly reusable collaboration scripts. *International Journal of Computer-Supported Collaborative Learning, 5*(3), 321-343.
Weinberger, A., Ertl, B., Fischer, F. & Mandl, H. (2005). Epistemic and social scripts in computer-supported collaborative learning. *Instructional Science, 33(1)*, 1-30.
Weinberger, A., Stegmann, K. & Fischer, F. (2010). Learning to argue online: Scripted groups surpass individuals (unscripted groups do not). *Computers in Human Behavior, 26*, 506-515.
Wood, D. J., Bruner, J. S. & Ross, G. (1976). The role of tutoring in problem-solving. *Journal of Child Psychology and Psychiatry, 17*, 89-100.

Acknowledgments
This Research was supported by a grant from the Deutsche Forschungsgemeinschaft (DFG).

Conversational Strategies that Support Idea Generation Productivity in Groups

Rohit Kumar, Language Technologies Institute, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA, rohitk@andrew.cmu.edu
Jack L. Beuth, Department of Mechanical Engineering, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA, beuth@andrew.cmu.edu
Carolyn P. Rosé, Language Technologies Institute, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA, cp3a@andrew.cmu.edu

Abstract: Our recent work has shown that conversational tutors that use social interaction strategies can achieve significant learning improvement through better management of student attention. In continuation with this line of investigation, this paper presents an experiment that measures the separate and joint effects of the nature and style of automatically generated interventions. We investigate aspects of designing intervention prompts from two different angles: first in terms of the addition of socially oriented turns over and above the instructional turns required to keep the students on track, and second in terms of the style of presentation of the tutor's instructional turns. We see positive effects of both manipulations, but on different outcome metrics. Thus, we conclude that the two directions can be productively combined into a single socially enhanced feedback approach, which contributes to the literature on dynamic forms of support for collaborative learning.

Introduction

Inquiry as an approach to learning is popular in many areas of instruction, but particularly in science and engineering instruction. *Idea generation* is of central importance in the inquiry process, and some prior work shows an important connection between idea generation productivity and learning in a scientific inquiry task (Wang et al., 2007). Notice that the goal for students here is not to select and then apply a known procedure for solving a problem. In contrast, the main purpose here is to let students actively and creatively generate the candidate problem solving steps/options by themselves. Beyond offering students the opportunity to generate possible solutions to problems, these tasks offer students the opportunity to weigh and balance trade-offs between alternative solutions proposed by themselves and their team mates. Thus, these problems draw on student creativity and offer opportunities for students to develop their social skills and creative problem solving skills at the same time.

While idea generation in groups is purported to be more effective than idea generation for individuals, it is a well known problem that when groups engage in idea generation together, a phenomenon referred to as *process loss* occurs. In particular, it has been repeatedly demonstrated that a group that is interacting while doing idea generation together may not always perform better than a collection of non-interacting individuals whose contributions are simply pooled afterwards (i.e., nominal groups), both in terms of the quantity and quality of unique ideas, and in fact may sometimes perform significantly worse (Hill, 1982; Diehl & Stroebe, 1987; Nijstad & Stroebe, 2006). Often inquiry learning tasks are done collaboratively in the classroom. To the extent that process losses that affect productivity in idea generation may lead to corresponding reductions in learning, the issue of process losses in group idea generation is an important issue to investigate. Nevertheless, learning in idea generation tasks may arise from multiple different mechanisms, not only from the idea generation process per se. For example, while evaluative statements may inhibit productivity in idea generation, it is a form of transactivity in collaborative discourse, which shows that group members are attending to one another's contributions and making explicit links between their contributions and those that came before, which is associated with learning (Joshi & Rosé, 2007). Supporting such behavior has been shown in other work to support learning (Weinberger et al., 2005). Thus, the mere existence of process losses does not lead us to conclude hastily that collaborative idea generation is ill advised from an instructional standpoint. Rather we seek to find support that can enable students to benefit from the interaction without incurring the negative side-effects.

Tutorial dialogue agents have been effective as an interactive form of support for collaborative learning including collaborative problem solving (Kumar et al., 2007a), collaborative design (Kumar & Rosé, 2011) and collaborative idea generation (Wang et al., 2007), leading to substantial improvements in learning sometimes greater than a standard deviation effect size (Kumar et al., 2007b). However, despite the effective support that automated tutors offer to students learning in groups, it has been reported that groups of students often ignore and abuse the tutor, unlike the case where students are individually tutored (Kumar et. al., 2007b, Bhatt et. al., 2004). We reason that the presence of other students in collaborative learning scenarios causes the tutors to compete for the attention of the students. Since the tutors do not participate in social interaction during the formative phase of the group, they are pushed to the periphery of the learning group and have to struggle to make themselves be heard. In our earlier work (Chaudhuri et. al., 2008, 2009), we have explored the use of interaction tactics that can help the tutor grab the conversational floor before presenting its instructional content to the students. On the other hand, empirical research in the area of small group communication suggests that leaders in small groups perform optimal levels of task related and socio-emotional interaction with the group members to maintain a non-peripheral position in the group (Bales, 1958). Existing research on conversational tutors has largely focused on presenting only task related information, i.e., lessons and instructions in the case of tutors. In this paper we report observations from a study in our continuing investigation on the effects that conversational agents in general can achieve if they are equipped with social conversational skills. Recent work explores how social strategies employed by conversational agents results in greater effectiveness of their support (Kumar et al., 2010; Ai et al., 2010).

Wang et al. (2007) demonstrated that automatically generated intervention prompts increased both idea generation productivity and learning, but did not fully compensate for the loss incurred by the group. In this paper, we focus specifically on the design of the prompts, carefully manipulating both its nature and its style in order to fine tune it for increased effectiveness. Our results point to a specific design inspired by work on analysis of social positioning from the systemic functional linguistics community as being particularly promising for enhancing idea generation productivity. Furthermore, the analysis of the connection between idea generation and learning in our data suggests that the relationship may be more complex than what was indicated by the prior Wang et al. study, which raises interesting questions for future work. We conclude by outlining our next steps, which build upon these results.

Designing Conversational Strategies

Theoretical and empirical study of group interaction processes has been of interest in the communications research community since the 1950's. McGrath (1984) reviews various theories that address the functions of group interaction processes. Among these, Robert F. Bales (1950) proposed that two fundamental processes operate within groups, i.e., instrumental (task related) vs. expressive (socio-emotional). Over attention on one of these processes causes lapses on the other. Hence, interaction shifts between these two in order to keep the group functional. Participations in the interaction balance their positioning among the task leader and social leader roles through the use of these two types of interaction processes.

In the case of conversational tutors, the task related interactions include common aspects like instructing students about the task, delivering appropriate interventions in suitable form and providing feedback. Work in the area of affective computing and its application to tutorial dialog has focused on identification of students' emotional states (D'Mello et. al., 2005) and using those to improve choice of task related behavior by tutors. However, there has been only limited study of expressive (socio-emotional) aspects of the tutor's conversations with learning groups (Kumar et al., 2010; Ai et al., 2010). Those studies demonstrate positive impacts of social strategies when conversational agents offer support to students learning in groups.

In our current work, we independently manipulate two different factors that influence how conversational agents deliver intervention prompts to students. One dimension is identical to that investigated in prior work (Kumar et al., 2010), which is based on Bales' IPA, and discussed in the next section. This factor determines whether or not certain social prompts are delivered to students in addition to task related prompts. The second dimension determines the form of the tutor's prompts. We apply the distinction between heteroglossic contributions and monoglossic contributions from Martin & White's Engagement system (Martin & White, 2005), where a heteroglossic assertion is made in such a way as to acknowledge the possibility of alternative perspectives and a monoglossic assertion is one that speaks matter-of-factly "as the voice of God", without such an acknowledgement. This distinction is discussed in greater detail below.

Social Behaviors Motivated by Bales' IPA

We have used the interaction process analysis (IPA) schema developed by Bales (1950) to identify the social behaviors that tutors can employ. Our choice of interaction categories from IPA is based on the appropriateness of the unit of analysis on which IPA is applied, i.e., individual utterances, since the tutor's behavior is realized typically one utterance at a time. IPA identifies three positive socio-emotional interaction categories: showing solidarity, precipitating tension release, and agreeing. We have mapped these categories to eleven practically implementable conversational strategies, which are distinguishable from each other and are relevant to interactive situation employed in our experiment. This mapping is shown in Table 1.

Most of these strategies are realized as prompts triggered at appropriate moments during the interaction by manually developed rules. For example, strategy 1e is triggered when one or more students in the group are found to be inactive for over 5 minutes. In this event, the tutor chooses to raise the status of the inactive students by eliciting contributions from them through a prompt like: *Do you have any suggestions Mike?*

Table 1: Social Interaction Strategies based on three of Bales' Socio-Emotional Interaction categories.

1. Showing Solidarity: *Raises other's status, gives help, reward*
1a. Do Introductions: *Introduce and ask names of all participants*
1b. Be Protective & Nurturing: *Discourage teasing*
1c. Give Reassurance: *When student is discontent, asking for help*
1d. Complement / Praise: *To acknowledge student contributions*
1e. Encourage: *When group or members are inactive*
1f. Conclude Socially
2. Precipitating Tension Release: *Jokes, laughs, shows satisfaction*
2a. Expression of feeling better: *After periods of tension, work pressure*
2b. Be cheerful
2c. Express enthusiasm, elation, satisfaction: *On completing significant task steps*
3. Agreeing: *Shows passive acceptance, understands, concurs, complies*
3a. Show attention: *To student ideas as encouragement*
3b. Show comprehension / approval: *To student opinions and orientations*

Prompt Variations Based on Martin and White's Engagement System

In Martin and White's (2005) specification of what counts as heteroglossic, three requirements must be met: First, some propositional content must be being asserted in some form, although it may be done in such a way as to communicate extreme uncertainty. All of the examples in Table 2 assert something, so they all meet this requirement. However, questions that are framed in such a way as the reader believes the speaker was asking an honest question, for which no specific answer seems to be presupposed do not count as heteroglossic. Rather than asserting something, questions request information. Interjections, like "Yay", that cannot be interpreted as ellipsis, and thus have no propositional content are not considered heteroglossic, however, fixed expressions like "no", and "yes" that implicitly assert the propositional content of the yes/no question they are a response to do count as expressing propositional content. For example, if someone asks "Can ease of use be increased by increasing the length of the wrench handle?" and if I say, "Yes.", then what I am asserting is "Ease of use can be increased by increasing the length of the wrench handle." The context makes it clear what I am asserting through ellipsis. Other forms of ellipsis (e.g., "titanium" in response to "Which type of metal would you choose?") and do-anaphora (i.e., "I did." In response to "Did you select a material for the wrench?") similarly also count as having propositional content.

Table 2: Examples of Heteroglossic and Monoglossic variations of the tutor's instructional turns.

Original/Neutral	*Think about how we can improve this?*
Heteroglossic	I suggest that you think about how to improve this
Monoglossic	Think about how to improve this
Original/Neutral	*Maybe we can use a safer material this time!*
Heteroglossic	Using a safer material would be strongly advisable!
Monoglossic	Use a safer material this time!
Original/Neutral	*We keep a factor of safety to avoid reaching yield stress.*
Heteroglossic	It would make sense to keep a factor of safety to avoid reaching yield stress.
Monoglossic	Keep a factor of safety to avoid reaching yield stress.

Second, an awareness must be made visible to the presence of alternative perspectives than that represented by the propositional content of an utterance. This distinction is illustrated through three examples that appear in Table 2. In each case we see a monoglossic and heteroglossic version for each of three propositions. In particular, bald claims, even if they are biased, do not acknowledge alternative perspectives. For example, "Titanium is the obviously superior choice" is undoubtedly subjective, but it is not heteroglossic. It doesn't show any awareness that someone else might disagree. If a speaker goes on, however, to give reasons to defend the statement, then that speaker is showing awareness of other perspectives. These cases will be caught by the third requirement, which is that in order to count as heteroglossic, the acknowledgement of other perspectives must be expressed grammatically (e.g., through a model auxiliary like "might") or paraphrastically (e.g., "I think") within the articulation of that propositional content. If it is implicit or signaled through the discourse structure, then that is not enough to count strictly as heteroglossic in the Martin and White sense, although they would acknowledge it as heteroglossic “in spirit”.

The Engagement system begins with the distinction between heteroglossia and monoglossia, which we have just discussed. Once we have determined that an utterance counts as heteroglossic, we can then further subdivide it into utterances that Contract the positions or perspectives that are treated as viable within a conversation, or conversely, ones that Expand the scope of what is treated as viable. Either way, an acknowledgement is made that more than one way of looking at the world is at play. Utterances that contract that scope, such as making an absolute assertion that leaves no room for questioning, or out right rejecting a position, are typically seen as taking a more authoritative stance than ones that expand the options, such as making a suggestion. This notion of levels of authoritativeness is one important component of expressing the positioning of the speaker in relation to the propositional content. However, it also says something about where the speaker positions himself in relation to the audience. Taking an authoritative stance casts the other speaker into a less authoritative stance. Furthermore, when this system is again further subdivided, we see other options for positioning. For example, what is referred to as a Distancing move, in which the source of authority is ascribed to a third party, allows an authoritative statement to be made, which may contract options, but does not interfere with the positioning between the speaker and the audience. The speaker remains committed to the authoritative proposition, but is not responsible for it.

Implementation of Conversational Agents

The interaction between the students and tutor in the experiment presented in this paper is situated in a freshmen engineering course. In this course college students learn about basic mechanical engineering concepts like force, moment, stress, etc. The students interact with an automated tutor as part of a computer-aided design competition where the students are asked to design a better wrench with consideration to ease of use, safety and material cost. Students could interact with each other and the tutor using a text-based chat room that includes a shared whiteboard (Stahl, 2006; Kumar & Rosé, 2011).

Table 3: Excerpt of a tutor providing a lesson to a team of four students.

	Speaker	Contribution
64	Tutor	Intuitively, if you wanted to make the wrench easier to use, would you make it longer or shorter?
67	S5	Longer
70	Tutor	That's right. A longer wrench is better.
72	Tutor	Why is a longer wrench easier to use though! Let's look at the concept of Moment.
73	Tutor	When you use a wrench to turn a bolt, do you want a higher or lower moment?
74	S16	Higher

The task of the tutor is to provide lessons on the underlying theoretical concepts while the students work through a worksheet to explore various design choices. An excerpt of a lesson on the concept about the relationship between the length of a wrench and its ease of use is shown in Table 3 above. Besides performing its task related functions, the tutor also employs the social interaction strategies listed in Table 1. We have implemented this tutor using the Basilica architecture. Details about the inner workings of the Basilica architecture are described in a separate publication (Kumar & Rosé, 2011).

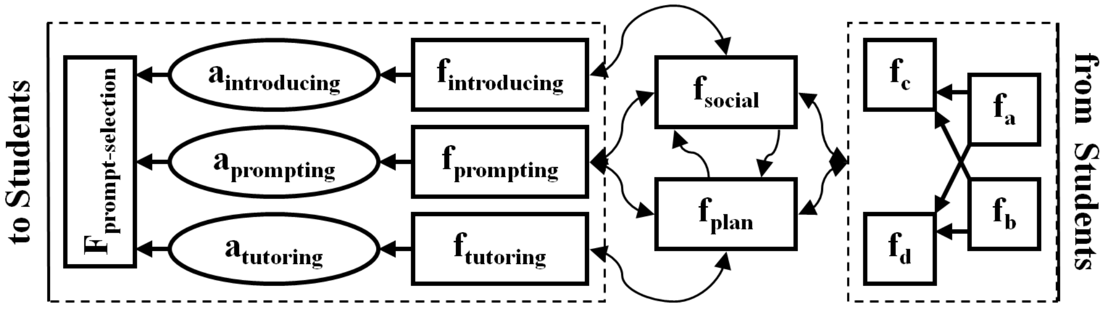


Figure 1. Basilica Component Network for Avis, the Freshmen Mechanical Engineering Tutor.

Using the Basilica architecture, conversational agents are modeled as a network of behavioral components. Each component implements a behavior that could be a combination of perception, thought and action. There are three types of components: actors (actuators / performers), filters (perceptors / annotators / coordinators) and memories. Each component can generate events carrying signals and data. Connected components can receive and process the events and generate further events. This architecture allows a developer to build agents by adding behavioral components incrementally. Since each component is only loosely coupled to a small number of other components, it provides the flexibility to easily change a single behavior. Also, it allows components to be reused between agent/tutor implementations for different tasks.

The tutor agent developed for the freshmen mechanical engineering learning domain is comprised of 22 Basilica components: four actors, sixteen filters and two memories. The tutor uses a gender neutral name Avis. Figure 1 below shows a simplified depiction of the component network of Avis. Actor components are shown as ovals and filter components are shown as rectangles. Arrows depict connections and possible directions of event flow.

Three of the actor and filter components correspond to three observable behaviors of the tutor, i.e., Introducing, Prompting/Hinting and Tutoring about concepts. Most of the other filters (depicted as f_a, f_b, f_c, f_d here) form a sub-network that processes student turns from the chatroom, as shown by filters f_a through f_d in Figure 1. This sub-network annotates turns with applicable semantic categories, accumulates them to identify inactive students and generates events that regulate the coordinators (f_{social} and f_{plan}).

The f_{plan} coordinator is responsible for executing the task-related interaction plan of the tutor. On the other hand, the f_{social} coordinator interleaves social prompts into the interaction. Before the prompts generated by the actor components are sent to the students, an appropriate variation of the prompt (*heteroglossic* / *monoglossic*) is selected by the prompt selection filter depending on the tutor's configuration.

Method

The rest of the paper describes the procedures and results from a controlled experiment we conducted to investigate the impact of the nature and style of feedback offered to students in an idea generation task.

Experimental Design

Our experimental design was a 2x3 between subjects factorial design. The first independent factor in our study manipulates whether social prompts motivated by Bales' IPA are inserted into the conversation. In the experimental condition (*Social*), students interacted with a tutor that was equipped with the eleven social interaction strategies, unlike the control condition (*Task*) which is our lower baseline condition. In both conditions, students would receive the same task related information (instructions / lessons / feedback) through the automated tutor. Based on the examples in Table 3 above, we notice that in the task condition, the tutor has features (like asking questions and giving feedback) that most common tutors do.

The second independent factor manipulates the form of the task related prompts presented by the tutor. In the first experimental condition (*Heteroglossic*), turns are presented in a Heteroglossic style. Approximately 90% of these were expressed in the Heteroglossic expand style, whereas about 10% were expressed in the Heteroglossic contract style. In the second experimental condition (*Monoglossic*), all task related prompts were authored in a Monoglossic style. In the Control condition (*Neutral*), employed as an ecological control, we used the same task prompts that were used in an earlier study (Kumar et al., 2010).

The time allotted for the interaction is the same for each group.

Procedure and Outcome Measures

We conducted a between-subjects experiment during a college freshmen computer-aided engineering lab project. 131 mechanical engineering students enrolled in the lab and participated in the experiment, which was held over six sessions spread evenly between two days. The two days of the experiment were separated by two weeks. Students were grouped into teams of three to four individuals. Each group communicated using a private chatroom (Stahl, 2006). No two members of the same group sat next to each other during the lab. The groups were evenly distributed between the six conditions in each session.

The procedure of the experiment comprised of 5 steps:

(1) Each session started with a follow along tutorial of computer-aided analysis where the students analyzed a wrench they had designed in a previous lab. Students spent about 25 minutes on the analysis.

(2) A pre-test with 11 questions (7 multiple choice questions and 4 brief explanation questions) was administered.

(3) The experimental manipulation happened during the Collaborative Design Competition after the pre-test. Students were asked to work as a team to design a better wrench taking three aspects into consideration: ease of use, material cost, and safety. Students were instructed to make three new designs and calculate success measures for each of the three aspects under consideration for their designs. They were also told that a tutor will help them with the first and the second designs so that they are well prepared to do the final design. No additional details about the tutor were given.

(4) After the students spent 35 minutes on the design competition, a post-test was administered.

(5) Following the test, student filled out a perception survey. The survey comprised of ten items to be rated on a seven point Likert-scale ranging from Strongly Disagree (1) to Strongly Agree (7). These questions

are shown in Table 4. Additionally, we asked students to check all of the following personality descriptors that they felt described the tutor agent: Encouraging, Inspiring, Assertive, Wishy Washy, Accommodating, Manipulative, Supportive, and Pushy.

Table 4: Items about Tutor and Learning Task rated by students on a 7-point Likert Scale.

Q1	The tutor provided very good ideas for the discussion.
Q2	I am happy with the discussion we had during the design challenge.
Q3	The tutor was part of my team.
Q4	The tutor was very cordial and friendly during the discussion.
Q5	The tutor received my ideas positively during the discussion.
Q6	The tutor helped keep tension low during the design activity.
Q7	The design challenge was exciting, and I did my best to come up with good designs.
Q8	Overall, I liked the tutor very much.
Q9	I think the tutor was as good as a human tutor.
Q10	Overall, I think we were successful at meeting the goals of the design challenge.

Results and Discussion

Idea Generation Productivity

In our analysis, the most important outcome factor is idea generation productivity. Idea generation productivity was estimated using an automatic measure developed in an earlier study using the same design task (Kumar et al., 2010). A dictionary was constructed using keywords from design ideas contributed in transcripts from the earlier studies. Counts of occurrences of words in this dictionary in the contributions of a student can serve as a rough estimate of idea generation productivity. Although this is a noisy measure, the effect of our experimental manipulation was strong enough to show significant differences between conditions. In order to identify the optimal design for feedback in an idea generation scenario from the standpoint of productivity, we used an ANOVA model with our estimate of idea generation productivity as the dependent variable. The two independent factors manipulated to determine the composition of feedback presented to students were each independent variables. We also included an interaction term that computes the interaction between the two independent factors in our design. There was no significant effect of the Social factor $F(1,125) = 0.0045$, $p = .97$. However there was a significant effect of the Heteroglossia manipulation $F(2,125) = 5.1$, $p < 0.01$, whereby the Heteroglossic condition was significantly higher than the Neutral condition (effect size 0.61 *s.d.*) and the Monoglossic condition (effect size 0.52 *s.d.*). Overall, this result confirms the hypothesis that a heteroglossic presentation style encourages more exploration and more productive idea generation.

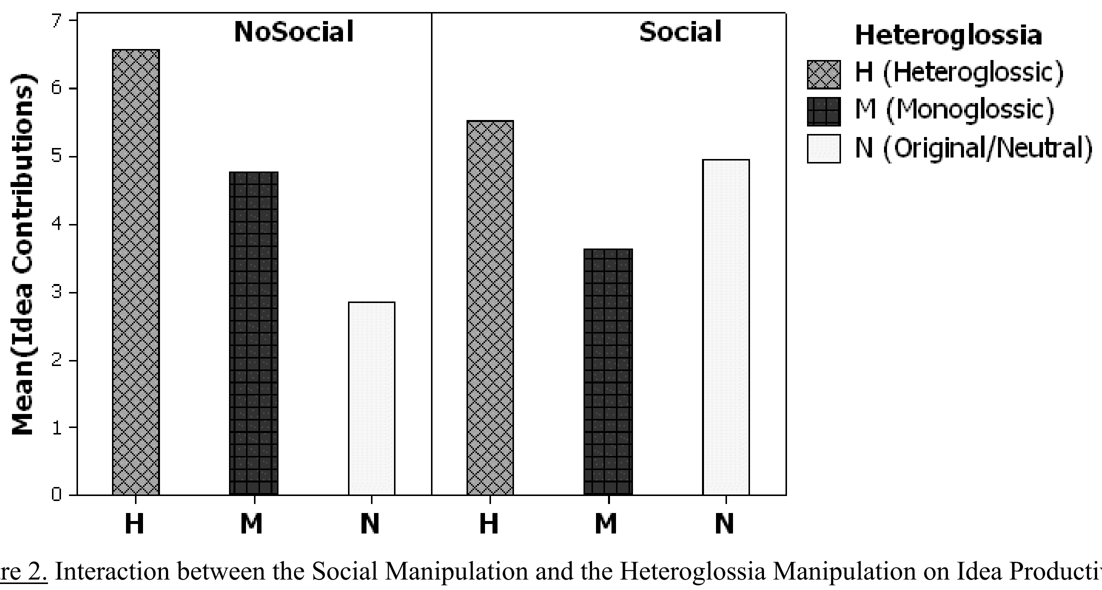


Figure 2. Interaction between the Social Manipulation and the Heteroglossia Manipulation on Idea Productivity.

There was a significant interaction effect (displayed in Figure 2) such that in the No Social condition, Heteroglossic is better than Neutral but not Monoglossic, and in the Social condition, Heteroglossic is better than Monoglossic, but not Neutral, $F(2,125) = 3.18$, $p < 0.05$. We find that the use of heteroglossic prompts could potentially compensate for the tutor's lack of social capabilities.

Learning Outcomes

There was no significant effect of either factor on learning, nor was there any significant interaction between factors on learning. Furthermore, whereas Wang and colleagues (2007) found a significant correlation between idea generation productivity and learning, we found no such relationship in our data. Wang and colleagues very carefully designed their preparatory reading materials, tests, and tasks in such a way that the act of generating an idea would provide opportunities for students to reflect on connections between facts presented in the reading materials and to generate bridging inferences akin to self-explanations (Chi et al., 1994). We believe this explains the difference in findings. In our task, learning occurs from deepening the understanding of principles through application. While we expect that factors that affect how deeply students reflect on the principles they are applying would have an effect on learning in our task, it is not necessarily the case that this level of engagement with the material corresponds to the number of design ideas students contribute to the discussion during the task. While we do not make any strong claims about learning and idea generation from our results, we believe this raises questions about the nature of the connection between learning and idea generation productivity, leading us to believe that the connection is complex and highly contextual.

Perception Ratings

On perceptual questionnaire, the Social manipulation had a relatively minimal effect. There were no significant main effects and no significant interactions between factors. There was a marginal effect of Social on Q5 such that students felt that the tutor received their ideas more positively in the Social condition, $F(1,125) = 3.55$, $p = 0.06$. Unfortunately, this is not necessarily a positive result since there was a negative correlation between the Q5 ratings and our idea productivity measure, $R = -0.34$, $p < 0.0001$.

The Heteroglossia manipulation had a larger effect. There was a significant effect of Heteroglossia on Q2, indicating that students in the Heteroglossia condition were significantly happier with the discussion than students in the Control condition, and the Monoglossic condition was not different from either, $F(2,125) = 3.65$, $p < 0.05$. There was also a significant effect of Heteroglossia on Q8, indicating that students like the Heteroglossic and Monoglossic tutor agents more than the agent in the Control condition, $F(2,125) = 3.48$, $p < 0.05$. There were also two marginal effects. First, students felt that the tutor agent was as good as a human tutor marginally more often in the Heteroglossic condition than the Control condition, with the Monoglossic agent not being different from either, $F(2,125) = 2.6$, $p = 0.08$. Similar to Q5, this was found to have a significant negative correlation with our idea productivity measure as well, $R = -0.22$, $p = 0.01$. Also, there was a marginal effect of Heteroglossia on Q10 such that students in the Heteroglossic condition felt that they were more successful than students in the Monoglossic condition, with students in the Control condition not being different from either, $F(2,125) = 2.84$, $p = 0.06$. There was no correlation between idea generation success and this perception of success.

Overall, these are relatively weak effects, although they do demonstrate some preference for the Heteroglossic style of presentation. However, when we evaluated the effect of condition on personality indicators using binary logistic regressions, the story became a little more complex. There were only significant effects of the Social manipulation such that the Social condition rated the tutor agent more Supportive and less Pushy than the Control condition. There was no effect of the Heteroglossic manipulation. Thus, we see that the Social manipulation had more of an effect on the perception of the tutor agent's personality whereas the Heteroglossia manipulation had more of an effect on how the students felt about working with the agent and how successful they were in their task. Since we do not see significant interactions between factors, we can conclude that we can achieve the best of both by combining the Heteroglossic style with the added social strategies of the Social condition.

Conclusion

First and foremost, the study presented in this paper shows that conversational tutors used in collaborative learning scenarios can be improved significantly by making them socially capable while keeping the task (tutoring) related behavior the same. In this study, we investigated social strategies from two angles. The first angle was to insert additional socially focused prompts over and above task related prompts. The effect of this addition was to make the tutor agent appear more supportive and less pushy. The second angle was to explore the style of presentation of task related prompts using the Martin and White (2005) Heteroglossia framework. Here we see an advantage in terms of supporting idea generation productivity in addition to eliciting more liking of the tutor agent and more satisfaction with the task.

In our future work we will continue to explore additional stylistic dimensions of tutor feedback and their associated effect on the student experiences in collaborative contexts, including their own social positioning within their groups, their effective participation within those groups, and ultimately their learning. In particular, we have discussed how the construct of heteroglossia introduces the notion that the voice of the speaker is situated among other voices. But beyond that acknowledgement of the existence of other voices, what we do not see in this simple binary distinction is the manner of that positioning. The details of that positioning are further specified within Martin and White's Appraisal framework (Martin & White, 2005), which includes Attitude, in which feelings are revealed towards propositional content, Graduation, in which feelings are either magnified or downlplayed, and Engagement, in which a speaker positions herself in relation to the propositional content of the utterance, positions the audience in relation to the propositional content, and positions herself in relation to the audience (Martin & White, 2005). The observed effects of the Heteroglossia manipulation in this study demonstrate how productive it may be to explore how constructs from sociolinguistics might inspire the design of more effective support for social interactions in computer supported collaborative learning.

References

Ai, H., Kumar, R., Nguyen, D., Nagasunder, A., Rose, C. P. (2010). Exploring the Effectiveness of Social Capabilities and Goal Alignment in Computer Supported Collaborative Learning, in *Proceedings of Intelligent Tutoring Systems,* Pittsburgh PA

Bales, R. (1958). Task Roles and Social Roles in Problem Solving Groups, *Readings in Social Psychology*, E.E. Maccoby, et al. (eds.), New York: Holt, Rinehart, and Winston

Bales, R.F. (1950). *Interaction process analysis: A method for the study of small groups*, Addison-Wesley, Cambridge, MA

Bhatt, K., Evens, M., Argamon, S. (2004). Hedged responses and expressions of affect in human/human and human/computer tutorial interactions. In *Proc. of the Cognitive Science Society*, 114-119, Chicago, IL.

Chaudhuri, S., Kumar, R., Joshi, M., Terrell, E., Higgs, F., Aleven, V., Rosé, C. P. (2008). It's Not Easy Being Green: Supporting Collaborative Green Design Learning. In *Proc. of Intelligent Tutoring Systems*, 807-809, Montreal, Canada.

Chaudhuri, S., Kumar, R., Howley, I., Rosé, C. P. (2009). Engaging Collaborative Learners with Helping Agents. In *Proc of AI in Education*, 365–372, Amsterdam.

Chi, M. H., De Leeuw, N., Chiu, M., Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18, 439-488.

Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: toward the solution of a riddle. *Journal of Personality and Social Psychology*. 53(3), 497-509.

D'Mello, S. K., Craig, S. D., Gholson, B., Frankin, S., Picard, R., Graesser, A. C. (2005). Integrating Affect Sensors in an Intelligent Tutoring System. In *Proc. of Workshop on Affective Interactions: The Computer in the Affective Loop at IUI*, 7-13, New York, NY

Hill, G. W. (1982). Group versus individual performance: are *N*+1 heads better than one? *Psychological Bulletin*, 91(3), 517-539.

Joshi, M. & Rosé, C. P. (2007). Using Transactivity in Conversation Summarization in Educational Dialog. *Proceedings of the SLaTE Workshop on Speech and Language Technology in Education*, 53-56, Farmington, PA.

Kumar, R., Gweon, G., Joshi, M., Cui, Y., Rosé, C. P. (2007a). Supporting students working together on Math with Social Dialogue. *Workshop on Speech and Language Technology in Education*, Farmington, PA.

Kumar, R., Rosé, C. P., Wang, Y. C., Joshi, M., Robinson, A. (2007b). Tutorial Dialogue as Adaptive Collaborative Learning Support. In *Proc. of AI in Education*, 383-390, Los Angeles, CA.

Kumar, R., Ai, H., Beuth, J., Rosé, C. P. (2010). Socially-capable Conversational Tutors can be Effective in Collaborative Learning Situations, in *Proceedings of Intelligent Tutoring Systems*, 156-164, Pittsburgh, PA.

Kumar, R. & Rosé, C. P. (2011). Architecture for building Conversational Agents that support Collaborative Learning, *IEEE Transactions on Learning Technologies Special issue on Intelligent and Innovative Support Systems for Computer Supported Collaborative Learning*, 4(1), 21-34.

Martin, J. R. & White, P. R. (2005). *The Language of Evaluation: Appraisal in English*, Palgrave

McGrath, J. E. (1984). *Groups: Interaction and Performance*. Prentice-Hall, NJ

Nijstad, B. A., & Stroebe, W. (2006). How the group affects the mind: a cognitive model of idea generation in groups. *Personality and Social Psychology Review*, 10(3), 186-213.

Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge*. Cambridge, MA: MIT Press.

Wang, H-C., Rosé, C.P., Cue, Y., Chang, C-Y. Huang, C-C., & Li, T-Y. (2007). Thinking hard together: the long and short of collaborative idea generation in scientific inquiry, in *Proceedings of Computer Supported Collaborative Learning*, New Brunswick, NJ.

Weinberger, A., Ertl, B., Fischer, F., & Mandl, H. (2005). Epistemic and social scripts in computer-supported collaborative learning. *Instructional Science, 33* (1), 1-30.

Acknowledgments

This research was supported in part by NSF EEC 0935145, NSF DRL 0835426, and NSF SBE 0836012.

The Myth of Over-scripting: Can Novices be Supported Too Much?

Karsten Stegmann, Jin Mu, Vera Gehlen-Baum, Frank Fischer, LMU München, Leopoldstrasse 13, 80802 München, GERMANY
Email: karsten.stegmann@psy.lmu.de, jin.mu@psy.lmu.de, vera.baum@psy.lmu.de, frank.fischer@psy.lmu.de

Abstract: Despite the fact that many researchers in CSCL use the term over-scripting to interpret negative effects of scripts, the term is not clearly defined. Our contribution is to reframe the term according to concepts of internal and external scripts. We further conceptualize potential interferences between internal and external scripts in terms of cognitive processes and motivation. In an empirical study (*N* = 81) we varied the degree of an argumentative script (low vs. medium vs. high) and examined the effects on processes and outcomes of argumentative knowledge construction. Our results show positive effects of the medium and high degree of scripting on argumentative knowledge construction. We found negative effects of the medium and high degree of scripting on motivation, however, but low motivation did not negatively interfere with knowledge acquisition. Furthermore, our reframing of over-scripting allowed us to differentiate between under-scripting, over-scripting and, finally, malfunctional scripts.

Origin of Over-scripting

In 2002, Pierre Dillenbourg wrote an influential paper with the title 'Over-scripting CSCL: The risks of blending collaborative learning with instructional design.' Since then, many authors have referred to the term over-scripting when they talk about the design and (negative, unexpected) effects of collaboration scripts (e.g. Beers et al., 2005; Karakostas & Demetriadis, 2009; Kollar et al., 2007). Over-scripting seems to be implicitly defined as a negative learning outcome when collaboration is scripted, rather than as the specific cognitive or motivational effect of scripts on collaborative processes. Thus, whereas the term over-scripting may help researchers to articulate their concerns regarding 'too much script', it is of doubtful value as a theoretical concept to guide research on the effects of collaboration scripts on collaborative knowledge construction. Using the term over-scripting in its current meaning to explain negative effects of scripts is therefore not of much help in the endeavour to understand better the mechanisms leading to positive or negative collaboration script effects.

Therefore, we will try to reframe the idea of 'too much script' regarding cognitive and motivational aspects in this contribution. Further, we aim to answer the question whether there can be too much collaboration script to support collaborating novices during online discussions of cases in an experimental study.

Reframing Over-scripting in Terms of Internal and External Scripts

Novices in terms of computer-supported collaborative learning can be regarded having a low-structured internal (collaboration) script (cf. Kollar et al., 2007). Schank and Abelson (1977) introduced the term to describe cognitive structures that represent context-dependent knowledge in sequences of activities and roles in specific situations (e.g., knowledge about sequence and roles during a school lesson). Scripts are abstract mental structures that organize the processing of sequences of events.

With respect to external collaboration scripts we follow the definition that they specify and sequence learning activities and/or assign roles to different learners (Kollar, Fischer, & Hesse, 2006) and can consequently facilitate specific discourse activities such as the construction of arguments or specific interaction patterns (e.g. Schellens et al., 2005; Weinberger et al., 2010). Kobbe and colleagues (2007) specified that the components of a collaboration script are *participants, activities, roles, resources,* and *groups*. The mechanisms of CSCL scripts that manipulate the components are *task distribution*, *group formation*, and *sequencing* (Kobbe et al., 2007).

Research on collaborative learning in online discussions has provided evidence that at least two argumentative internal scripts are related to learning: epistemic activities and providing warrants for claims. A central epistemic activity is the application of conceptual knowledge on problem information (cf. Weinberger & Fischer, 2006). The higher the amount of the application of conceptual knowledge, the higher the epistemic quality of argumentation and the higher the knowledge acquisition is. Providing warrants for the relation between grounds and claims has been regarded as a process very similar to self-explanations (cf. Baker, 2003). The more claims and grounds are connected with a warrant, the higher the quality of argumentation.

The next two sections will now focus on problems that may occur if external and internal scripts interfere. We identified two main areas where internal and external scripts may interfere and potentially undermine collaborative knowledge construction: First, suboptimal fit of internal and external scripts may have negative effects on cognitive processes during collaboration. Second, external regulation may have negative effects on motivation of learners during collaboration.

Potential Interference between Internal and External Scripts on Cognitive Processes

The perfect fit between external and internal collaboration would be if existing (but not activated) internal scripts are activated by an external script and enough information is provided regarding all activities that are not yet part of the internal script of a learner. This perfect fit is, however, the ideal rather than the typical case. It is more likely that an external script addresses scenes (e.g. 'Please provide a counterargument!') without sufficient information on the activity or addresses scripts that are not known (e.g. 'Please engage in knowledge building by integrating pros and cons!'). The domain-specific knowledge acquisition as well as the internalization of the script is not supported. This kind of interference between internal and external scripts might be called 'cognitive under-scripting'.

If an external script provides scaffolds that guide procedures for which internal scripts are already represented by the learner or where a learner might even hold more effective or efficient internal script, the performance of the learner will decrease. Instead of performing highly automated cognitive skills, learners slowdown because they have to follow the external script. The time-in-task is reduced, whereas the time on handling the script is unnecessarily increased. The internalization of the script is restricted to scripts that were not held before. The acquisition of domain-specific knowledge is not facilitated owing to the lowered efficiency. This kind of interference between internal and external scripts might be called 'cognitive over-scripting'.

If collaboration lasts a certain amount of time, the originally perfect fit of internal and external script may change towards cognitive over-scripting owing to an ongoing internalization of the script (cf. Wecker & Fischer, 2009). To avoid this cognitive over-scripting, the script might be faded out, i.e. components like examples and explanations vanish or several small steps collapse to bigger steps. A script may ask the learner to provide a complete argument instead of asking for claim, ground and warrant (e.g. Kollar et al., 2007).

Last but not least, an external collaboration script asks the learner to perform activities that do not facilitate collaborative knowledge construction. Such an external script may even have a positive effect on the internalization of the script, but the domain-specific knowledge acquisition might be hindered. At this point we have to clarify that a script might lead to a 'better' online discussion, but not to better learning processes. The negative effect of the script on domain-specific knowledge acquisition compared with an unsupported control condition should *not* be regarded as 'over-scripting'. This kind of interference between internal and external scripts produces a malfunctional scripting. The script is functioning, but with negative effects on collaborative knowledge construction.

The interference between internal and external scripts depends on the degree to which the (internal or external) script specifies activities. The more details of activities are specified the higher the degree of scripting.

Potential Interferences between Internal and External Scripts on Motivation

Motivation is an important factor in learning. Motivation is usually regarded as affecting the effort a learner is willing to invest, i.e. to what extent a learner contributes to collaboration. If instruction reduces motivation it carries the risk that engagement in collaborative knowledge construction will decrease. Even the best (internal or external) script does not foster learning if learners are not motivated to activate or use it, respectively.

Deci and Ryan (1991) posited that a decrease of the feeling of autonomy may cause a loss of intrinsic motivation. External collaboration scripts usually reduce the degree of freedom of learners and lower their perceived autonomy. Therefore, learners may put less effort into activity or strategy guided by the script. As a consequence, processes and outcomes of collaborative knowledge construction might be negatively affected. This effect might occur independently of the internal script of a learner, but is related to the extent of how autonomy is reduced. Against this background, scripts that scaffold procedures on the low level of activities should always have negative effects on intrinsic motivation. The reduction of autonomy and thereby intrinsic motivation is then mainly script immanent. The goal, however, has to be to realize scripts that are as 'motivation friendly' as possible.

Research Questions

We are investigating the effects of degree of scripting on processes and outcomes of argumentative knowledge construction in online discussions. The goal is to examine whether a non-malfunctional script can cause over-scripting when novice (regarding argumentative knowledge construction) are supported. The first research question examines, therefore, the central perquisite, that the scripted activites are positively related to domain-specific knowledge acquisition.

RQ1: To what extent are epistemic quality and quality of argumentation related to acquisition of domain-specific knowledge? According to the approach of argumentative knowledge acquisition we assume that both, the epistemic quality as well as the quality of argumentation is positively related to domain-specific knowledge acquisition.

RQ2: To what extent does the degree of scripting (without/low vs. medium vs. high) have on effect on the quality of argumentative knowledge construction? If the collaboration script is not malfunctional, it should have positive effects on the epistemic quality and/or quality of argumentation, but not decrease the quality of argumentative knowledge construction.

RQ3: To what extent does the degree of scripting (without/low vs. medium vs. high) have an effect on motivation during argumentative knowledge construction? According to our reframing of motivational interferences between external and internal scripts, the self-reported intrinsic motivation as well as the objectively measured amount of contributions to online discussion during collaboration might be negatively affected by the degree of scripting. The higher the degree of scripting, the lower the intrinsic motivation and amount of contributions should be.

RQ4: To what extent has the degree of scripting (without/low vs. medium vs. high) effects on the acquisition of domain-specific knowledge and knowledge of argumentation? The degree of scripting is expected to have a positive effect on domain-specific knowledge as well as knowledge of argumentation. If, however, the medium degree of scripting out-performs the high degree scripting condition, cognitive over-scripting occurrs (if RQ1 and RQ2 provided evidence that the script is not malfunctional).

RQ5: To what extent does quality of argumentative knowledge construction and motivation during argumentative knowledge construction predict the acquisition of domain-specific knowledge? This research question finally examines, which of the aspects that might be negatively affected by interferences between external and internal scripts has the strongest influence on domain-specific knowledge acquisition. The answer to this question may allow us to weigh potential negative effects on motivation against positive effects on specific scripts.

Methods

Participants and Design

Eighty-one (81) students of Educational Science at the University of Munich participated in this study during the summer term 2010. The mean age of the participants was $M = 23.36$ ($SD = 3.85$) years. Participation was a requirement for receiving course credit in a mandatory introductory course for undergraduates because the experimental learning environment was part of the regular curriculum. We manipulated the variable 'degree of scripting' (without/low vs. medium vs. high) by means of a computer-supported collaboration script for argumentative knowledge construction that is described below. The participants were randomly assigned to groups of three. The groups were then randomly assigned to one of the three experimental conditions in the one-factorial design.

Learning Environment and Implementation of Collaboration Scripts

The subject of the learning environment was Weiner's attribution theory (1985) and its application in education. The students read the text of this theory and the text of introducing argumentation individually before the experimental session. In the present study, three problem cases from practical contexts were used as a basis for online discussions. The group's task was to analyze the three cases and to come up with a joint solution for each case. The three students in each group were distributed within a laboratory room, i.e. they were together in one room, but not sitting next to each other. An asynchronous, text-based discussion board was used for collaboration. This discussion board allowed the exchange of text messages that resembled emails.

The computer-based learning environment used in this experiment is a modified version of the one employed by Stegmann, Weinberger and Fischer (2007). All of the instructions are presented in the form of standard videos which vary according to the treatments. The collaboration scripts for supporting argumentative knowledge construction were implemented in a tool called S-COL (Scripting for Collaborative Online Learning) which allows for the sustainable development of scripts and scaffolds that can be used with a broad variety of content and platforms (Wecker et al., 2010).

(1) The script in the condition with a high degree of scripting consisted of several components. Each learner was responsible for the analysis of one of the three cases. The collaboration phase was divided into ten phases. Each learner was asked to provide an analysis for her or his case in the first phase. During the next phases two and three, each learner had to provide counterarguments for the first analysis of her or his learning partners. During the fourth phase, the learner had to integrate the counterarguments of her or his learning partners into the analysis. Phases two to four were repeated twice. During the tenth phase, however, the learners were tasked with writing the final analysis for their case. The progression from phase to phase was automatized, i.e. after a certain time, learners were automatically forwarded to the next phase. In each phase the script specified several steps. The script guided learners first to analyse the case or the arguments of the learning partners and second to construct (counter)arguments. For example, the script for the construction of single arguments asked learners to select a person, acting in the case, to analyse it, to specify a concept from theory that was applicable to the subject, to formulate a claim using the subject and the theoretical concept, to provide a ground (i.e. case information) that supported the claim and finally to provide a warrant that explained why the ground supported the claim. For each step of the script the system provided an explanation. For each step of the

script an example was provided. Despite the fact the script specified these steps, however, the learners were not forced to follow them. They always had the chance to write 'unscripted' statements in the message box.

(2) The medium degree of scripting was realized through fading-out of the steps during the different phases as well as the fading-out of the explanations and examples. We decided us against a more or less randomly selected degree of scripting between low and high, because that would have meant that learners received different amounts of information regarding the script. Through implementing the medium degree of scripting using fading, we are able to provide subjects wihtin in this condition the very same information on the script, but have different degree of scripting on average.

(3) The condition with low degree of scripting received no additional support in solving the three problem cases, i.e. the condition can be regarded as without script. All participants, however, also in this condition, were advised to argue well according to the text on argumentation they had to read before the experimental session.

Procedure

First, the participants completed pre-tests that were designed to measure domain-specific prior knowledge and prior knowledge of argumentation. The data from these tests were used to control randomization. All instructions were presented in the form of videos to standardize the procedure. The groups collaborated for 80 minutes, trying to develop analyses for the three cases and to reach agreement about them. In the final phase (about 35 minutes), the students took individual post-tests on domain-specific knowledge and knowledge of argumentation. Time-on-task was held constant for the three conditions.

Data Sources

Online discussions of five problem cases (one pretest case, three cases during collaboration, and one posttest case) served as data sources. The online discussions were coded in terms of epistemic quality and quality of argumentation. Furthermore, amount of contributions was measured on the basis of the online discussion on the three cases during collaboration. Intrinsic motivation and knowledge on argumentation were measured using an online questionnaire after the online discussion of the post-test case.

We used individual sentences or parts of a compound sentence as the unit of analysis (segment) (cf. Strijbos et al., 2005). Each sentence was coded whether it was a claim (i.e., a statement that advances the position learners take to analyse case with attribution theory), a ground (i.e., evidence from case to support claim), a warrant (i.e. logical connections between the grounds and claims that present the theoretical reason why a claim is valid) or something else (i.e. all other sentences).

Two human coders analysed almost one tenth of the raw discourse corpora (distributed over five cases), which have been further used for training the customized algorithms for automatic coding on the multiple categories by SIDE tool (Mayfield & Rosé, 2010). We achieved a high value of Cohen's Kappa = .96 for segmentation, and the inter-rater agreement between two human coders to classify the text was Cohen's Kappa = .71. A human coder and SIDE achieved an agreement of Cohen's Kappa = .97 (accuracy = 99.3%) to segment the conversational data into the units of analysis. The objectivity across all cases comparing SIDE with a human coder was sufficiently high (Cohen's Kappa = .81; accuracy = 84.5%).

Subsequently, all claims were analysed regarding epistemic activities, i.e. the correctness of application of theoretical concepts on case information. The median of reliability between two human coders was sufficiently high with Cohen's Kappa value = .89.

Dependent Variables

Prior domain-specific knowledge. Appropriately applied concepts from the Weiner's attribution theory in the context of analysing the pre-test case were used to measure prior domain-specific knowledge.

Prior knowledge of argumentation. The proportion of sentences in the online discussion on the pre-test case that was coded as warrant was used to measure the application of argumentative knowledge during online discussion. By using the proportion instead of the total amount, the value is controlled for the influence of domain-specific knowledge.

Amount of contributions in pre-test discussion. The amount of contributions during pretest discussion was measured by the number of sentences written by a learner during the online discussions of the pre-test case.

Intrinsic motivation. The instrument used to investigate the motivational processes that occur during learning was a questionnaire developed by Prenzel and colleagues (1993). The reliability of the scale was sufficiently high (Cronbach's alpha = .73).

Amount of contributions during collaboration. The amount of contributions during pretest discussion was measured by the number of sentences written by a learner during the online discussions of the three cases.

Epistemic quality. The number of appropriately applied concepts from Weiner's attribution theory in the context of analysing the cases *during collaboration* was used to measure the epistemic quality.

Quality of argumentation. The quality of argumentation during online discussion was measured in the same way as the prior knowledge on argumentation, but using the online discussion during the collaborative learning phase as data source.

Domain-specific knowledge. The number of appropriately applied concepts from Weiner's attribution theory in the context of analysing the post-test case was used to measure domain-specific knowledge.

Knowledge of argumentation. Participants were asked to name the components of a formally complete single argument and to formulate a single argument on the topic of smoking. One point was given for each of the three correctly applied components, i.e. claim, ground and warrant. Hence, the test scores could range from zero to six points. Two trained coders evaluated the tests independently with a sufficiently high reliability (Cohen's Kappa = .94).

Application of argumentative knowledge. The proportion of sentences in the online discussion on the post-test case that were coded as warrant was used to measure the application of argumentative knowledge.

Statistical Tests

To examine the effects of the degree of scripting, we used a linear regression analysis with the dummy variables script (without vs. with) and degree of scripting (medium vs. high). This method allowed us to examine effects of the script in general and effects of the degree of scripting in particular with only one statistical analysis. Therefore, it is more efficient than an ONEWAY ANOVA with two subsequent analyses of contrasts. To control for differences prior to the treatments within the different conditions and no effects of specific scripting degrees are expected, we compared the three different conditions using an ONEWAY ANOVA.

Results

Preliminary Data Analyses

First, we controlled for differences between conditions prior to the intervention by applying an ONEWAY ANOVA. No significant effects of the degree of script regarding epistemic quality, $F_{(2, 78)} = 0.11$, $p = .90$, $\eta < .01$, quality of argumentation, $F_{(2, 78)} = 2.15$, $p = .12$, $\eta^2 = .05$, or amount of contributions, $F_{(2, 78)} = 0.18$, $p = .833$, $\eta^2 < .01$, in the online discussion on the pre-test case were found.

RQ1: Relation between Quality of Argumentation and Domain Specific Knowledge

The first research question is intended to provide evidence that the processes that we are going to foster with the collaboration script are related to individual knowledge acquisition. For both aspects, epistemic quality and quality of argumentation, partial correlation with domain-specific knowledge in the post-test was computed to control for prior knowledge.

Significant partial correlation have been found between epistemic quality and knowledge acquisition, $r_{par} = .26$, $df = 69$, $p = .01$, one-tailed, as well as regarding quality of argumentation, $r_{par} = .29$, $df = 69$, $p = .01$. Our assumption that the quality of argumentative knowledge construction during online discussion is associated with the acquisition of domain-specific knowledge is supported by this result.

Table 1: Epistemic quality, quality of argumentation, intrinsic motivation and amount of contributions during online discussion by experimental condition: means (m) and standard deviations (SD).

		Degree of scripting		
		Without script/low	medium	high
Epistemic quality	*M*	13.24	10.52	10.47
	SD	12.00	8.78	7.00
Quality of argumentation	*M*	8.75	11.77	15.88
	SD	6.11	6.91	8.80
Intrinsic motivation	*M*	2.53	2.03	1.94
	SD	0.90	0.65	0.62
Amount of contributions	*M*	57.65	33.50	24.88
	SD	20.60	11.60	12.39

RQ2: Quality of Argumentative Knowledge Construction

The effects of the degree of scripting on the quality of argumentative knowledge construction will be examined regarding two different aspects: epistemic quality and quality of argumentation.

Epistemic quality. The regression model regarding epistemic quality was not significant, $R^2 = .02$, $F_{(2,78)} = 0.66$, $p = .52$. Neither the factor script, $\beta_{stand.} = -.14$, $p = .29$, nor the factor degree of scripting, $\beta_{stand.} < .01$, $p =$

.98, were significant predictors of the epistemic quality during the online discussion. The script did not facilitate the epistemic quality of the online discussion (see table 1 for descriptive values).

Quality of argumentation. The model using the predictors script and degree of scripting explained about 12% of the variance of the quality of argumentation, $R^2 = .12$, $F_{(2, 78)} = 5.25$, $p = .01$. The factor script had a significant positive effect on the quality of argumentation, $\beta_{stand.} = .37$, $p < .01$. Learners supported by a script (medium as well as high degree of scripting) had a higher quality of argumentation during online discussion than learners without script. The factor degree of scripting had a marginally positive effect (significant at the 10%-level) on the quality of argumentation, $\beta_{stand.} = .22$, $p = .07$. Learners with the medium degree of scripting benefited less regarding the quality of argumentation than learners with the high degree of scripting (see table 1 for descriptive values).

RQ3: Motivation during Argumentative Knowledge Construction

The effects of the degree of scripting was examined with respect to two different measures: intrinsic motivation and amount of contributions.

Intrinsic motivation. The regression model regarding intrinsic motivation was significant, $R^2 = .107$, $F_{(2, 78)} = 4.57$, $p = .01$. The factor script had a significant negative effect on intrinsic motivation, $\beta_{stand.} = -.35$, $p < .01$. Learners supported by a script (medium as well as high degree of scripting) showed lower intrinsic motivation than learners without script. The factor degree of scripting had no effect on intrinsic motivation, $\beta_{stand.} = .06$, $p = .60$. The fading of the script neither positively nor negatively affected intrinsic motivation of learners (see table 1 for descriptive values.

Amount of contributions. The regression model regarding amount of contributions was significant, $R^2 = .42$, $F_{(2, 78)} = 28.30$, $p < .001$. The factor script had a significant negative effect on amount of contributions, $\beta_{stand.} = -.71$, $p < .001$. Learners supported by a script (medium as well as high degree of script) contributed less sentences to the online discussion than learners without script. The factor degree of scripting had a positive effect on amount of contributions, $\beta_{stand.} = .22$, $p = .03$. The medium degree of script diminished the negative effect of the script on amount of contributions, i.e. learners with a medium degree of script contributed more sentences to the online discussion than learners with the high degree of script (see table 1 for descriptive values).

Post hoc, the question was, to what extent the effects of script and degree of scripting on amount of contributions are mediated by intrinsic motivation. We have to clarify whether this negative effect on amount of contributions through a loss of motivation or not. Therefore, we conducted a mediator analysis, i.e. we computed the residual of the model where intrinsic motivation predict amount of contributions and checked whether the factors script and degree of scripting still had an effect on the residual. If the effect disappears or the effect size is much smaller than on the initial variable, mediation has taken place. If no mediation can be found, other, non-motivational, factors, may have mediated the effect of script and degree of scripting on motivation.

Mediator analysis. Predicting amount of contributions with intrinsic motivation resulted in a significant model, $R^2 = .13$, $F_{(2, 78)} = 11.52$, $p = .001$. Intrinsic motivation is a significant positive predictor of amount of contributions, $\beta_{stand.} = .36$, $p = .001$. The higher the intrinsic motivation of learners, the higher the number of sentences they contributed to the online discussion. The factors script and degree of scripting, however, are still significant predictors of the residual, $R^2 = .36$, $F_{(2, 78)} = 21.05$, $p < .001$. The explained variance regarding the residual is 15% lower than regarding amount of contributions directly. Thus, intrinsic motivation partly mediates the effect of script and degree of scripting on amount of contributions, but a major part of the variance cannot be explained by motivation.

RQ4: Acquisition of Domain-specific Knowledge and Knowledge of Argumentation

The effects of the degree of scripting on knowledge acquisition were examined regarding three different aspects: domain specific knowledge, declarative knowledge of argumentation, and script internalization.

Domain specific knowledge. No effect of the degree of scripting on domain-specific knowledge could be found, $R^2 = .02$, $F_{(2, 78)} = 0.76$, $p = .47$. Neither the factor script, $\beta_{stand.} = .12$, $p = .38$, nor the factor degree of scripting, $\beta_{stand.} = .05$, $p = .68$, were significant predictors of domain-specific knowledge in the post-test (see table 2 for descriptive values).

Knowledge of argumentation. The model using the factors script and degree of scripting to predict the knowledge of argumentation was significant, $R^2 = .20$, $F_{(2, 78)} = 9.53$, $p < .001$. The factor script had a significant positive effect on declarative knowledge of argumentation, $\beta_{stand.} = .34$, $p = .003$. Learners supported by a script (medium as well as high degree of scripting) were better able to list components of single arguments according to Toulmin (1958) and to provide a complete argument on smoking than learners without script. The factor degree of scripting had no significant effect on declarative knowledge of argumentation, $\beta_{stand.} = .18$, $p = .12$ (see table 2 for descriptive values).

Application of argumentative knowledge. While the overall model predicting the application of argumentative knowledge using the factors script and degree of scripting failed to reach significance, $R^2 = .05$,

$F_{(2, 78)}$ = 1.91, p = .16, the factor script had a marginal significant positive effect, $\beta_{stand.}$ = .24, p = .05. The factor degree of scripting, $\beta_{stand.}$ = -.12, p = .33, was not a significant predictor of the internal script on argumentation. Learners supported by the script seemingly developed a better internal script. The fading of the script in the condition with a medium degree of scripting did not interfere with this effect (see table 2 for descriptive values).

Table 2: Domain-specific knowledge, knowledge of argumentation and application of argumentative knowledge by experimental condition: means (m) and standard deviations (SD).

		Degree of scripting		
		Without script/low	medium	high
Domain-specific knowledge	*M*	4.62	5.76	5.40
	SD	2.91	3.28	3.07
Knowledge on argumentation	*M*	3.14	4.90	4.20
	SD	1.80	0.89	1.19
Application of argumentative knowledge	*M*	9.64	12.91	15.64
	SD	9.89	11.04	11.64

RQ5: Predicting Domain-specific Knowledge with Quality of Argumentative Knowledge Construction and Motivation

This research question aimed to answer which of the aforementioned aspects of collaborative learning in online discussions, i.e. intrinsic motivation, amount of contributions, epistemic quality and quality of argumentation are most relevant for the acquisition of domain-specific knowledge. Therefore, we tested a regression model for domain specific-knowledge in the posttest with the predictors intrinsic motivation, amount of contributions, epistemic quality, and quality of argumentation. The overall model was significant, R^2 = .15, $F_{(2, 78)}$ = 2.82, p = .03. Aspects of quality of argumentative knowledge construction, epistemic quality, $\beta_{stand.}$ = .27, p = .03, as well as quality of argumentation, $\beta_{stand.}$ = .27, p = .04, were significant positive predictors of domain-specific knowledge in the posttest. The more inferences between case information and theoretical concepts are drawn and the higher the proportion of warrants, the higher is the domain-specific knowledge in the posttest. Intrinsic motivation, $\beta_{stand.}$ = .12, p = .35, as well as amount of contributions, $\beta_{stand.}$ = -.06, p = .66, *were not* significant predictors of domain-specific knowledge in the posttest.

Discussion

Our results provide evidence for the basic assumptions of argumentative knowledge construction. Both epistemic quality and quality of argumentation are positively related to domain-specific knowledge acquisition. Furthermore, we have shown that our collaboration script fostered the quality of argumentation without impairing epistemic quality during online discussions.

As expected, the script affected intrinsic motivation negatively. A lower degree of scripting could, however, not reduce this effect. The perceived over-scriptedness seems rather dichotomous than continuous. Furthermore, the script had a negative effect on amount of contributions. This effect was less strong for the medium degree scripting than for the high degree scripting. A mediator analysis revealed that the negative effect of the script on the amount of contributions could not fully be explained by the loss of motivation. Instead, it seems to be more probable that learners supported by a script spent more time with learning processes that did not lead directly to more contributions to the online discussion, e.g. deeper elaboration (Stegmann et al., in press).

The script had a positive effect on both knowledge of argumentation and the internal script on argumentation. Learners supported by the script had more explicit knowledge of argumentation in a knowledge test on argumentation and, much more importantly, applied the script more frequently in an unsupported online discussion after the learning phase. Findings by Kollar and colleagues (2007) showed that learners with highly structured internal argumentation script outperformed learners with low structured internal scripts on argumentation regarding domain-specific knowledge acquisition. It can be argued that the script might have a sustainable effect, because the internalised script might facilitate future collaborative learning.

A positive effect of the script, however, could not be found with respect to domain-specific knowledge immediatley after the collaboration. In line with the argumentation on amount of contributions, we argue that this can be seen as a trade-of function: given constant time, learners who elaborate single concepts in more depth do have less time to elaborate a broad range of concepts and vice versa. The regression analysis showed that, the epistemic quality as well as the quality of argumentation is a significant predictor of domain-specific knowledge gains. Motivation and amount of contributions are less predictive here. These results show that for predicting learning outcomes, learning activities are better than motivational aspects. Intrinsic motivation can be helpful for collaborative learning, but it is not a prerequisite. In particular, novices to collaborative learning need to acquire internal scripts for successful collaboration. This process can be supported through collaboration scripts. In summary, this study did not provide any evidence of cognitive over-scripting of novices using computer-supported collaboration scripts.

Finally, the reframing of over-scripting allows differentiation between under-scripting, over-scripting, and malfunctional scripting. Our conceptualization allows us to classify scripts using result patterns of the effects of the script on process and outcomes as well as the relations between processes and outcomes. This is a substantial improvement compared with the conceptualization of over-scripting so far. Our approach will not be sufficient, however, for those researchers who are describing over-scripting *only* as “disturbing ‘natural’ interactions and ‘natural’ problem solving processes”. These problems are problems of ‘too much script’, but only if learners have access to functional internal scripts on collaboration that can be disturbed. In particular novices on collaboration have not. Therefore, negative effects of scripts on outcomes or processes should not any longer simply be classified as over-scripting. Instead, it has to be shown that the externally provided script is not just a malfunctional script.

References

Baker, M. (2003). Computer-mediated argumentative interactions for the co-elaboration of scientiWc notions. In J. Andriessen, M. Baker, & D. Suthers (Eds.), *Arguing to learn: confronting cognitions in computer-supported collaborative learning environments* (Vol. 1, pp. 1–25). Dordrecht: Kluwer.

Beers, P. J., Boshuizen, H. P. A., Kirschner, P. A., & Gijselaers, W. H. (2005). Computer support for knowledge construction in collaborative learning environments. *Computers in Human Behavior*, *21*, 623-643.

Black, A. E. & Deci, E. L. (2000). The effects of instructors’ autonomy support and students’ autonmous motivation on learning organic chemistry: A self-determination theory perspective. *Science Education*, *84*(6), 740-756.

Deci, E.L. & Ryan, R.M. (1991). A motivational approach to the self: integration in personality. In R.A. Dienstbier (Ed.). *Perspectives on motivation. Nebraska Symposium on Motivation*, (Vol. 38, 237-88). Lincoln: University of Nebraska Press.

Dillenbourg, P. (2002). Over-scripting CSCL. In P. A. Kirschner (Ed.), *Three worlds of CSCL: Can we support CSCL* (pp. 61-91). Heerlen: Open University of the Netherlands.

Karakostas, A., Demetriadis, S. (2009). Adaptation patterns in systems for scripted collaboration. Paper presented at *CSCL 2009*. International Society of the Learning Sciences (ISLS).

Kobbe, L., Weinberger, A., Dillenbourg, P., Harrer, A., Hämäläinen, R., Häkkinen, P., & Fischer, F. (2007). Specifying Computer-Supported Collaboration Scripts. *International Journal of Computer-Supported Collaborative Learning*, *2*, 211-224.

Kollar, I., Fischer, F., & Hesse, F. W. (2006). Collaboration scripts - a conceptual analysis. *Educational Psychology Review*, *18*(2), 159-185.

Kollar, I., Fischer, F., & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning & Instruction*, *17*(6), 708-721.

Mayfield, E. & Rosé, C. P. (2010). An Interactive Tool for Supporting Error Analysis for Text Mining. In *Proceedings of the North American Chapter of the Association for Computational Linguistics*

Prenzel, M., Eitel, F., Holzbach, R., Schoenheinz, R. J. & Schweiberer, L. (1993). Lernmotivation im studentischen Unterricht in der Chirurgie. *Zeitschrift für Pädagogische Psychologie*, *7*(2/3), 125-137.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding*. Hillsdale, NJ: Erlbaum.

Schellens, T., Van Keer, H., & Valcke, M. (2005). The impact of role assignment on knowledge construction in asynchronous discussion groups: A multilevel analysis. *Small Group Research*, *36*, 704-745.

Stegmann, K., Wecker, C., Weinberger, A., & Fischer, F. (in press). Collaborative Argumentation and Cognitive Elaboration in a Computer-Supported Collaborative Learning Environment. *Instructional Science*.

Strijbos, J-W, Martens, R. L., Prins, F. J., Jochems, W.M.G. (2006). Content analysis: What are they talking about? *Computers & Education*, *46*, 29–48.

Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge Universty Press.

Wecker, C., & Fischer, F. (2009). Preventing relapses into novice strategies during the fading of instructional scripts: The role of distributed control. Paper presented at *EARLI 2009*, August 25 to 29, 2009.

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computersupported collaborative learning. *Computers & Education*, *46*(1), 71-95.

Weinberger, A., Stegmann, K., & Fischer, F. (2010). Learning to argue online: Scripted groups surpass individuals (unscripted groups do not). *Computers in Human Behavior*, *26*, 506–515.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, *92*, 548-573.

A Scaffolded Software Tool for L2 Vocabulary Learning: GroupScribbles with Graphic Organizers

Wenli Chen, Chee-Kit Looi, Yun Wen, Learning Sciences Lab, National Institute of Education, Nanyang Technological University, 1 Nanyang Walk, Singapore 637616
Email: wenli.chen@nie.edu.sg, cheekit.looi@nie.edu.sg, yun.wen@nie.edu.sg

Abstract: Understanding technology from the perspective of the scaffolding can help us bridge the gap between abstract or general CSCL design principles and the design and enactment of concrete CSCL practices. This study addresses this issue by describing and discussing how we use GroupScribbles (GS) technology coupled with appropriate pedagogical graphic organizers to scaffold effective collaborative learning in the context of L2 students' vocabulary learning in Singapore classrooms. It is found that the GS technology and pedagogical graphic organizers can jointly scaffold students' collaborative vocabulary learning to achieve desired learning outcomes. When equipped with graphic organizers to help students plan and organize their problem solving, GS is transformed from a general tool for enabling seamless interactions to a scaffolded software tool integrated with pedagogical design for supporting specific learning, by problematizing important disciplinary content.

Introduction

There has been a discursive shift from knowledge transmission to knowledge construction in second language (L2) learning literature (Warschauer, 1997; Warschauer & Kern, 2000) since the late 1990s. The integration of collaborative learning into L2 learning has been emphasized (e.g., Belz, 2002; Lee, 2004) concomitant with the shift which regards knowledge as not something eternal or unchanging or that exists apart from humans but something that is the product of human activity. Learners are able to ultimately enhance their lexical ability through generating, sharing and improving their conceptual artifacts (e.g. grammatical rules or meaning of words) by interacting with each other. Learners can improve both syntactic ability as well as lexical ability through the collaboration with others.

This shift of the epistemological perspective prompts much needed changes in traditional classrooms. One such change is to seek in a more seamless way the integration and orchestration of individual learning, small-group and whole-class interactions in the classroom. Various technologies are used in classrooms to enhance the teacher-student and student-student interactions with the belief that technologies promotes productive meaningful interaction for L2 learner's (e.g., Belz, & Kinginger, 2002; Lan, Sung, & Chang, 2007). The scaffolding metaphor has been used to describe additional forms of support and contexts for enabling interaction. How to design or implement technology to scaffold learners' language learning through CSCL has long been a challenge for both researchers and teacher facilitators. This is particularly true in real classroom learning where the context for CSCL is complex and dynamic. Understanding technology from the perspective of the scaffolding can help us bridge the possible gap between abstract or general design principles and enactment of concrete CSCL practices. This study describes and discusses how we use GroupScribbles technology equipped with pedagogical graphic organizers as a scaffolded software tool to enhance L2 students' vocabulary learning in Singapore classrooms.

Literature Review

Scaffolded Software Tools

Scaffolding situations are those in which the learner gets assistance or support to perform a task beyond his or her own reach if pursued independently when "unassisted" (Wood, Brunet & Ross,1976, p. 90). This is related to Vygotsky's concept of Zone of Proximal Development (ZPD, defined as the zone of activity in which a person can produce with assistance what they cannot produce alone (or can only produce with difficulty). Traditional views of scaffolding focused on the social dimension - interactions with other people (teachers or peers) as the source of assistance articulating how a more knowledgeable person can help (Hogan & Pressley, 1997; Wood et al., 1976). In the past decade, many researchers started to investigate the scaffolding functions of the technological tools (Davis & Linn, 2000; Edelson et al., 1999; Guzdial, 1994; Quintana et al., 1999; Reiser et al., 2001). The technology dimension of scaffolding examines how technology changes the task in some way so that learners can accomplish tasks that would otherwise be out of their reach. The technology scaffolding include providing prompts to encourage or remind students what steps to take (Davis & Linn, 2000), graphical organizers or other notations to help students plan and organize their problem solving (Quintana et al., 1999), or representations that help learners track what steps they have taken (Collins & Brown, 1988; Koedinger & Anderson, 1993).

Pea (2004) summarized the 2 general properties for how scaffolding functions for the learner: 1) channeling and focusing - reducing the degrees of freedom for the task at hand by providing constraints that increase the likelihood of the learner's effective action; recruiting and focusing attention of the learner by marking relevant task features, with the result of maintaining directedness of the learner's activity toward task achievement, and 2) modeling: modeling more advanced solutions to the task. Reiser (2004) explained the "mechanisms" that a technology provides scaffolding for learners by differentiating between *task structuring* ("guiding learners through key components and supporting their planning and performance" [p. 273]) and *content problematizing* ("tools can shape students' performance and understanding of the task in terms of key disciplinary content and strategies and thus problematize this important content" [p. 273]).

The social dimension and the technological dimension of scaffolding is complementary – they work together in helping learners learn better in the sense that certain scaffolding activities can be the responsibility of the people (teacher or peers) and other scaffolding activities provided by the technology (Reiser, 2004). In this way, a scaffolding synergy (Tabak, 2004) could be achieved in support of student learning. This paper focuses on the technology dimension of the scaffolding for L2 vocabulary learning.

CSCL and L2 Language Learning

Collaborative learning has been considered to be one of the effective instructional strategies in language learning (Cohen, 1994; Dagenais, & Walsh, 2008; Wen, Looi, & Chen, 2011). It "has a 'social constructivist' philosophical base, which views learning as construction of knowledge within a social context and which therefore encourages acculturation of individuals into a learning community" (Oxford, 1997, p 443). The learners co-construct knowledge by working on joint problems or tasks, including making individual contributions, partaking in discussion and arriving at joint solutions. In the literature on Second Language Acquisition, learners' active involvement in interaction with other learners of the target language has been identified as a fundamental aspect of the learning process as it provides opportunities for students to engage in negotiation of meaning (Gutierrez, 2003). In doing so, they adopt ideas from their peers and share specific conceptions after collaborating. They influence one another, and converge or diverge with respect to language knowledge.

When learners learn together, members of the group are becoming similar with respect to their knowledge, as knowledge equivalence and as shared knowledge prior to, during, and subsequent to collaborative learning (Weinberger *et al.* 2007). This will lead to knowledge convergence, which is an increase in common knowledge that all collaborating learners had (Jeong & Chi, 2007). Learners who converge in knowledge have been found to benefit more from collaborating than learners who do not (Fischer & Mandl, 2005).

Computer technologies play an important role in supporting students' collaborative learning. Networked technologies offer the potential to bring collaborative learning to new heights. In a CSCL environment, participants are actively and collaboratively engaged in creating knowledge, and the collaboration is taking place through a computer network They engage in a coordinated effort to perform a task together to establish common knowledge (Littleton & Häkkinen,1999).

The lexicon may be the most important component for language learners (Gass & Selinker, 2001). Vocabulary is recognized a fundamental to the development of L2 proficiency (Harley, 1996). The vocabulary learning activities through CSCL approach can be classified into 2 categories: incidental and intentional vocabulary learning. "Incidental learning is said to occur when a word is acquired as a by-product of the learner being engaged in some language learning activity" (Smith, 2004, p.368). There is fairly solid evidence supporting the view that computer-assisted collaborative reading and writing benefits students' vocabulary acquisition, in which environment they have less anxiety and more opportunities to negotiate around anew or problematic words (e.g, Ghaith, 2003). Intentional vocabulary learning is task-based which focuses on vocabulary learning. Most classroom vocabulary learning activities are intentional learning activities. Existing CSCL research in language learning focuses more on incidental learning than intentional learning. This paper discusses the intentional vocabulary learning in a F2F (face-to-face) CSCL environment in real classrooms.

Background of the Study

Chinese as L2 in Singapore

Singapore is a multiethnic and multilingual country, in which bilingualism refers to proficiency in English which is recognized as the L1, and the L2 known as a "Mother Tongue Language" (MTL). Chinese or Mandarin is Chinese students' MTL. The Ministry of Education in Singapore defines MTL not by the language used at home or the first language learned by the student but by ethnicity (Tan, 2006). Hence, one of the characteristics of L2 learning in Singapore is the difference of students' language abilities in a class due to their different MTL backgrounds. For typical ethnic Chinese students in Singapore, their lack of motivation and limited basic linguistic capability continue to be the fundamental challenges to their learning in Chinese Language (Sim, 2005). Our school-based research is to address the typical linguistic weaknesses of younger Chinese in their L2 proficiency especially their command of vocabulary.

Chinese is a character-based language. To be literate in Chinese, one needs to know over a thousand Chinese characters (Wing, et al., 2003). From the perspective of linguists, the Chinese script, due to its logographic nature, is considered the most difficult script to learn by non-native learners (Shen, 2004). Shen (2002) attributes the challenge to the retention of the combination of the three elements of a character, that is, its sound, shape and meaning in learner's long-term memory, and the instant retrieval of these three elements. Innovations in language education have been targeted towards a more comprehensive understanding of the development of children's capability in handling the script of Chinese, and ways of enhancing learners' structural understanding of the writing system beyond rote learning and mechanical practice (Tse, 2002).

Research Design and Participants

We introduced CSCL in F2F Chinese language classrooms to help the students learn better by collaboratively constructing language knowledge together through authentic verbal discussion (in Chinese) mediated and enhanced by GroupScribbles. A design-research approach is adopted to address complex problems in real classroom contexts in collaboration with practitioners, and to integrate design principles with technological affordances to render plausible solutions (Brown, 1992; Collins, 1992). In the 3 year school-based research, we have worked with one primary school and 3 secondary schools in Singapore in systematically designing and implementing collaborative learning supported by GroupScribbles for Maths, Chinese, and English language learning (see Chen & Looi, 2010; Looi & Chen, 2010) . The research reported in this paper focuses on Chinese language learning, and specifically vocabulary learning.

All the GS-based collaborative activities were co-designed by the teachers and researchers. The students work on groups of four in their collaborative discussions. The students in a group are seated next to each other with a Tablet PC, and they engage in face-to-face talk. The Chinese language learning curriculum in Singapore Primary schools focuses on vocabulary (Liu & Zhao, 2008) whereas the Chinese language curriculum in secondary schools emphasizes textual or discourse structure and grammar. To be aligned with the specific lesson objectives as stated in the syllabus, our GS-based activities for primary school students were designed to assist students to learn new characters and word formation. In secondary schools, the GS activities on vocabulary were designed to help students review characters and words that they had learnt before. This paper will describe two activity designs for vocabulary learning at the primary and secondary school levels respectively.

GroupScribbles (GS)

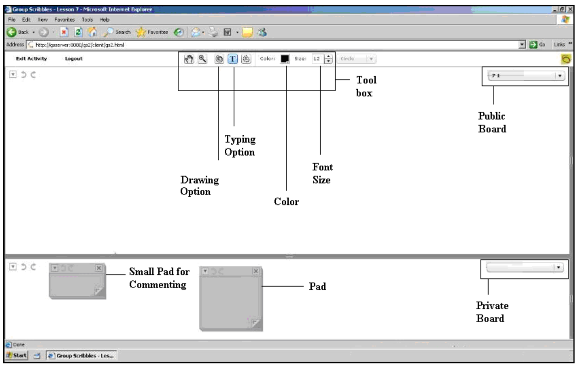


Figure 1. The User Interface of GS with a Two-Paned Window.

The technology tool we used to support students' collaborative learning is GroupScribbles (GS), which was co-developed by SRI international and National Institute of Education Singapore. The GS user interface presents each user with a two-paned window. The lower pane is the user's personal work area, or 'private board', with a virtual pad of fresh 'scribble sheets' on which the user can draw or type (see Figure 1). When the scribble sheet is moved to the public board, it can be synchronized to all the Tablet PC's public board so that all the members can view it. The essential feature of the GS client is the combination of the private board (where students can work individually) and group boards or public boards (where students can post their work and position them relative to the work of others, view others' work and take items back to the private board for further elaboration. It enables collaborative generation, collection and aggregation of ideas through a shared space based upon individual effort and social sharing of notes in graphical and textual forms. In our school-based research, GS is used routinely in F2F classroom setting. Effectively supporting students' F2F interactions as well as coordination in collaborative learning becomes possible with the use of GS.

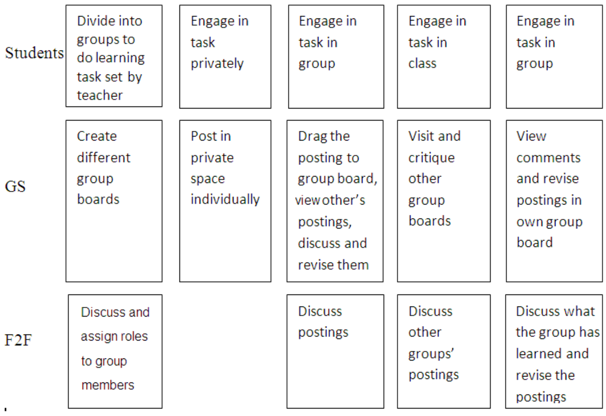


Figure 2. A Generalized Student Activity Pattern in GS Classroom.

Figure 2 shows a generalized student activity pattern supported by GS which we have found to be effective for productive language learning. One particular set of affordances of the GS-based learning environment is the ability for students to engage in private interactions in the GS private board, group interactions in the group board together with F2F verbal discussion, and class-level interactions through GS. When students work in the GS private space without others being able to see or directly impact them, they can engage with their materials and sense-making processes individually in a focused way (Vahey, Tatar, & Roschelle, 2007). We believe that it is very important to allow the student to complete as much of the vocabulary task as possible individually. When they drag the posting to the group board, they engage in the group interaction by constructing and discussing about the product or their work (Vahey et al., 2007). Then they visit other group boards to engage in class interactions where they are exposed with full range of ideas and artifacts, and they critique these ideas/artifacts. At the end of this activity, they go back to their own group board to engage in group or private interactions where they improve their ideas/artifacts after seeing other groups ideas/artifacts and comments. It is evident that GS technology scaffolds the process of different levels of interactions and the seamless switch between them: private interaction - group interaction - class interaction – group/private interaction. It enables a synergy between autonomy and collaboration by combining both private and collaborative learning. The F2F GS environment leverages resources such as shared screen, gestures, and conversation norms to help students jointly construct meaning, become more proficient in participating in representation-based interactions, and build a common understanding of the subject matter (Chen, Looi, & Tan, 2010; Vahey et al., 2007).

Graphic Organizers

It is not the technology that is the innovation in student learning, but the integration of technology and pedagogical practices in scaffolding the meaningful vocabulary leaning activities. Here we present 2 pedagogical graphic organizers in GS that help students' vocabulary learning. Figure 3 displays an organizer for guiding primary school students to learning new words/characters. The purpose of this graphic organizer is to enhance students' awareness of the character components, and to help students learn the usage of the words in authentic context by composing sentences that make use of the words. Different groups were responsible for different new words. They could learn individual new word through visiting one another's group board. Group students first were required to write down in the centre of the GS board the character/phrase they work with. In this case it was a phrase "不理不睬" (meaning "to completely ignore"). The space surrounding the phrase was divided into 7 sections, each of which was dedicated for one task. Starting from the top right in clockwise order, the tasks were: 1) to explain the meaning of the phase; 2) to give a few similar characters as the last character of the phrase; 3) to give a few homophones of the last character of the phrase; 4) to use the last character of the phrase to form words (word formation); 5) comments from other groups (similarly this group was required to give comments on other groups' work); 6) to expand the phrase into a sentence, and 7) to collocate the phrase with other words (word collocations).

Figure 4 shows a group board with the organizer completed by one of the groups. With the sub-tasks in different themes, all group members have an opportunity to contribute their individual knowledge and ideas to the group, no matter what language proficiency they have. Students who are not good at Chinese language can post the meaning of the phrase/word with the help of a dictionary. Students with higher Chinese language capability can take the responsibility for making sentences. All students can contribute homophones and similar characters for the given characters based on their different prior knowledge of the characters. Within the group, the task is not dominated by one or two students. Rather, all the students play their part to complete the task. At the class level, when the teacher asks students to visit other group boards to learn other words and give their comments to help others refine their group products, students with different Chinese language proficiency have an equal opportunity to give and receive constructive comments. In Figure 4, we can see the constructive comments (as circled out) from other groups which help them correct the wrongly formed Chinese characters.

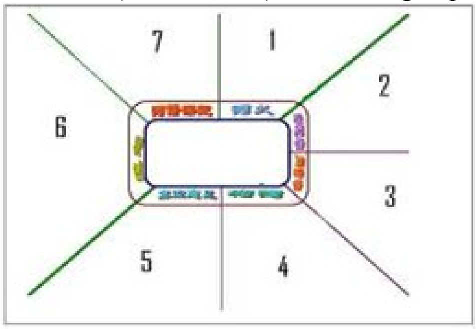


Figure 3. A Graphic Organizer for New Character & Word Learning.

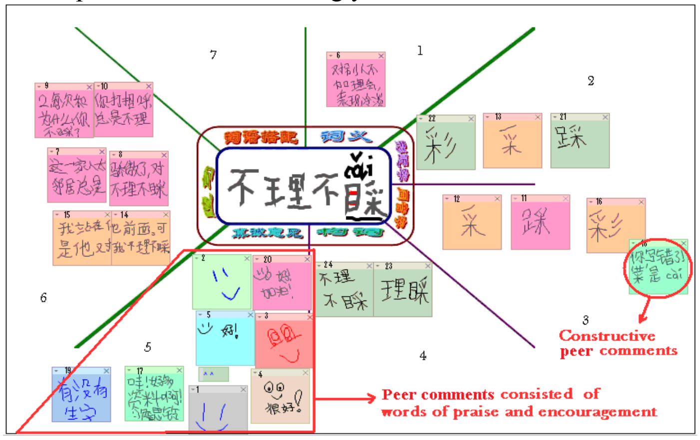


Figure 4. A Completed Group Board with Organizer (Primary School Example).

Next we describe the graphic organizer for the vocabulary learning of secondary school students. The Chinese script system is a principled- and rule-based system. Each Chinese character is comprised of one or more components, spatially arranged with certain principles (Liang, 2004). There are different standards to classify the basic spatial configurations for individual characters. In this activity we adopt a regular categorization, by which all the Chinese characters can be classified into 5 groups. They are up-down, up-mid-down; left-right; left-mid-right and full/half enclosed structure.

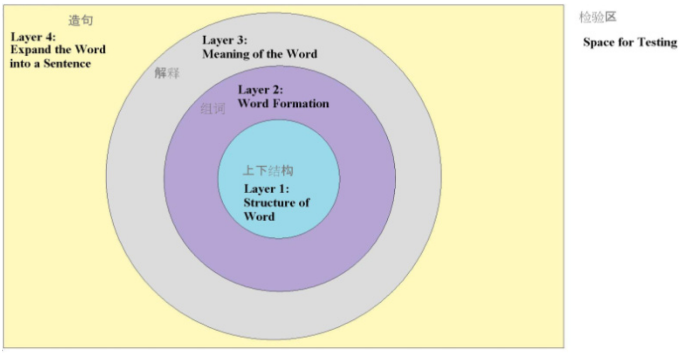


Figure 5. Graphic Organizer for New Character & Word Learning (Secondary School Example).

Figure 5 shows an organizer for helping students review vocabulary that they have learnt in the past semester. The centre of the figure is a solid circle with the instruction "上下结构" (meaning "up-down structure"). Students need to post characters with up-down structure onto this area. The three outer layers, working outward, are layer 2 "word formation"; layer 3 "meaning of the word"; layer 4 "expand the word into a sentence". Students are required to use the posted characters to form words; to give their own explanation for the words and to make proper sentence using the words. During this process, the students in a group can share their individual vocabulary knowledge and get a better understanding of the use of the words by negotiating their meanings in different context. Some blank space of the figure is left for assembling comments from the students from other groups.

Round Robin is suggested to be used together with this organizer in activity design, with the aim of knowledge sharing and construction during the whole class. Different groups are responsible for different tasks in the beginning. For example (see the protocol showed in Figure 6), group 1 students are responsible for picking up characters with up-down structures from the words list, and posting them onto the central circle, group 2 students are responsible for those up-mid-down structured characters, and so on. After all the groups complete their task in round 1, followed the protocol of round robin, group 1's students go to group 2's board to post related words onto the task of word formation. At this moment, group 2's students are working at group 3's board on left-right structured new characters. In round 3, group 1's students enter into group 3's public board for explaining the meaning of the words. A whole activity completes until all the groups finish checking the propriety of the postings from all other groups in round 5. Figure 7 displays the screenshot of a group board with artefacts.

Round / Group	Round 1: Pool Characters	Round 2: Form words	Round 3: Express meanings	Round 4: Create sentences	Round 5: Do a test
1	Public board 1（up-down）	Public board 2	Public board 3	Public board 4	Public board 5
2	Public board 2（up-mid-down）	Public board 3	Public board 4	Public board 5	Public board 1
3	Public board 3（left-right）	Public board 4	Public board 5	Public board 1	Public board 2
4	Public board 4（left-mid-right）	Public board 5	Public board 1	Public board 2	Public board 3
5	Public board 5（full/half enclosed）	Public board 1	Public board 2	Public board 3	Public board 4

Figure 6. Protocol of Round Robin.

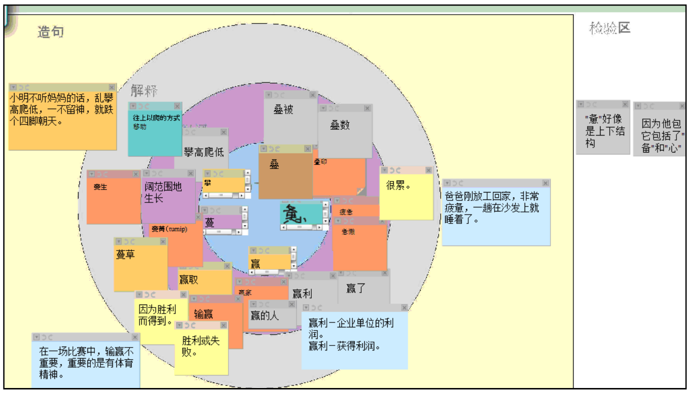


Figure 7. A Completed Group Board with Organizer for Reviewing Vocabulary.

Conclusion and Discussion

The teaching of L2 to young students, which involves complex linguistic skills and is thus cognitively demanding, has always been a great challenge to language teachers and researchers. The collaborative learning activity scaffolded by GS technology and graphic organizers was developed with the pragmatic aim of addressing the fundamental linguistic difficulties and challenges of younger students in learning L2 vocabulary. As Collins (1997) argued, "learning difficulties reflect differences, not deficiencies" (p. 3). Collaborative learning can help address the learning difficulties because students are more motivated to help each other when they work together, and they feel less threatened when they made mistakes, as their group-mates will help them. Consequently, they achieve improvement as a group together. GS activities help address the different language ability of students. It is hoped that this phenomenon of ability differentiation will be gradually faded out as all students will overcome their respective weaknesses, and therefore could contribute to the collaborative vocabulary learning. In these 2 learning scenarios, the GS technology and graphic organizers play important roles in scaffolding students to collaboratively learn vocabulary better.

From the perspective of Reiser's (2004) framework of scaffolding mechanisms, GS technology provides scaffolding for learners more by *task structuring* whereas graphic organizers serve as scaffolding by *content problematizing*. The design of GS technology supports the seamless switch between different types of cognition (private cognition – group cognition – class cognition – group cognition – private cognition") during the collaborative learning process. It guides students through the key collaboration procedures and supports their vocabulary learning planning and performance. Each individual student is not only responsible for his individual learning, but also for the group learning and class learning. Students must make their vocabulary knowledge public through GS that represent diverse knowledge explicitly. The artifacts created by students in GS become a vehicle for negotiation of understanding about the vocabulary knowledge. The shared representation can serve as a catalyst for negotiation of ideas.

With the graphic organizer, the students were provided with a pedagogical structure to scaffold their vocabulary learning. The mechanism of scaffolding by graphic organizers serves more as content problematizing by "shaping students' performance and understanding of the task in terms of key disciplinary content and strategies and thus problematizing this important content" (Reiser, 2004, p. 273). Chinese vocabulary learning is complex. There are different ways in learn such a character-based language. Traditionally, the students learn the characters by listening to a teacher's lecturing which focuses on one or two components (e.g., phonetics, structure, radicals) of the characters. This approach is not systematic and ignores the use of the characters or words in authentic contexts. The graphic organizers were designed to alleviate the problem by changing the nature of the vocabulary learning task through highlighting the key disciplinary content and strategies for vocabulary learning. They can scaffold vocabulary learning by "channelling and focusing" (Pea, 2004). In the GS activity, the organizers help to reduce the degrees of freedom for the task by providing constraints that increase the likelihood of the learner's effective action (focusing on all key components of vocabulary learning such as the meaning, structure, homophones, similar characters, uses in authentic sentences etc). The graphic organizers help recruit and focus attention of the learner on the relevant task features so that students are guided on a track toward task achievement. They help learners to perform by potentially enabling them to focus on more productive parts of the tasks. In our interview with the teacher, she expressed her view: "[with the graphic organisers] as time goes on, not only can students' interest of Chinese characters learning improve, they can internalize these strategies to learn Chinese vocabulary. They can also apply these strategies when they come across a new character or word. After some time, the students may no longer need these tangible scaffolding organizers anymore because they have learned what to do when they learn other new characters or words". Thus, the teacher's view is consistent with the expertise reversal effect (Kalyuga, Ayres, Chandler, & Sweller, 2003).

Scaffolding is distributed across the tools and context in which learning is happening. GS technology and graphic organizers jointly scaffold students' collaborative vocabulary learning to achieve the desired learning outcomes with technological scaffolding synergistically provided by the GS technology and by the pedagogical graphic organizers. The GS technology scaffolds students' collaboration and distributed cognition occurring across activities, artifacts, within group and across groups. The GS is a shared workspace for all the students to share knowledge and ideas. Students are able to view the artefacts generated by other students. They provide feedback to other groups' artefacts, question each other and argue about the meanings, which in turn build community knowledge. Therefore, in this F2F CSCL environment, vocabulary knowledge is not an object that is acquired and possessed by individuals, but becomes embedded in the conversations and social discourse. Without the graphic organizers, GS is a general collaboration tool that supports different level of interactions by taking in to account both individual and social processes in CSCL. It scaffolds students in both knowledge acquisition and knowledge as participation (Sfard, 1998). When equipped with graphic organizers to help students plan and organize their problem solving, GS is transformed from a general tool for enabling seamless interactions to a scaffolded software tool integrated with pedagogical design for supporting specific learning goals by problematizing important disciplinary content. Therefore by having both GS tool and appropriate graphic organizers, scaffolding synergy could be achieved in support of the learner's advances by allowing greater opportunities for vocabulary learning that otherwise would not be possible in the classroom.

References

Belz, J. A. (2002). Social dimension of telecollaborative foreign language study. *Language Learning & Technology, 6*(1), 60-81.

Belz, J. A., & Kinginger, C. (2002). "The cross-linguistic development of address form use in telecollaborative language learning: Two case studies." The Canadian Modern Language Review 59(2): 189-214.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions. *Journal of the Learning Sciences, 2*, 141–178.

Chen, W., & Looi, C.K. (2010). What do Students do in a F2F CSCL Classroom? The Optimization of Multiple Communications Modes. *Computers & Education, 55*(3), 1159-1170.

Chen, W., & Looi,C.K. (In press). Active Classroom Participation in a GroupScribbles Primary Science Classroom. *British Journal of Educational Technology, 41*.

Chen, W. & Looi, C.K. (2010). What do Students do in a F2F CSCL Classroom? The Optimization of Multiple Communications Modes, *Computers & Education*, *55*(3), pp. 1159-1170.

Cohen, E. (1994). *Designing group work* (2nd). New York: Teachers College Press.

Collins, A. (1992). Toward a design science of education. In E. Scanlon&T. O'Shea (Eds.), *New directions in educational technology* (pp. 15–22). Berlin, Germany: Springer-Verlag.

Collins, A., & Brown, J. S. (1988). The computer as a tool for learning through reflection. In H. Mandl & A. Lesgold (Eds.), *Learning issues for intelligent tutoring systems* (pp. 1–18). New York: Springer-Verlag.

Collins, J.L. (1997). *Strategies for Struggling Writers*. Guilford Press.

Dagenais, D., & Walsh, N. (2008). Collaboration and co-construction of knowledge during language awareness activities in Canadian elementary school. *Language Awareness, 17*(2), 139-155.

Davis, E. A., & Linn, M. C. (2000). Scaffolding students' knowledge integration: Prompts for reflection in KIE. *International Journal of Science Education, 22,* 819–837.

Edelson, D. C., Gordin, D. N., & Pea, R. D. (1999). Addressing the challenges of inquiry-based learning through technology and curriculum design. *The Journal of the Learning Sciences, 8*, 391–450.

Gass, S. M., & Selinker, L. (2001). Second language acquisition. Mahwah, NJ: Erlbaum.

Ghaith, G. (2003). Effects of the learning together model of cooperative learning on English as a foreign language reading achievement, academic self-esteem, and feelings of school alienation. *Bilingual Research Journal, 27*(3), 451-475.

Gutierrez, G. A. G. (2003). Beyond interaction: the study of collaborative activity in computer-mediated tasks. *ReCALL, 15*(1), 94-112.

Guzdial, M. (1994). Software-realized scaffolding to facilitate programming for science learning. *Interactive Learning Environments, 4*, 1–44.

Harley, B. (1996). Vocabulary learning and teaching in a second language. *Canadian Modern Language Review, 53,* 3-12.

Hogan, K., & Pressley, M. (1997). Scaffolding scientific competencies within classroom communities of inquiry. In K. Hogan & M. Pressley (Eds.), *Scaffolding student learning: Instructional approaches and issues* (pp. 74–107). Cambridge, MA: Brookline.

Jeong, H., & Chi, M. T. H. (2007). Knowledge convergence and collaborative learning. *Instructional Science, 35*, 287-315.

Fischer, F., & Mandl, H. (2005). Knowledge convergence in computer-supported collaborative learning: The role of external representation tools. *The Journal of the Learning Sciences, 14*(3), 405-441.

Kalyuga, S. Ayres, P., Chandler, P. & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist,* 38(1) 23–31.

Koedinger, K. R., & Anderson, J. R. (1993). Reifying implicit planning in geometry: Guidelines for model-based intelligent tutoring system design. In S. P. Lajoie & S. J. Derry (Eds.), *Computers as cognitive tools* (pp. 15–45). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Lan, Y. J., Sung, Y. T., & Chang, K. E. (2007). A mobile-device-supported peer-assisted learning system for collaborative early EFL reading. *Language Learning & Technology, 11*(3): 130-151.
Lee, L. (2004). Native and nonnative speakers' perceptions of networked collaborative interaction: A sociocultural approach. In L. Lomicka & J. Cooke-Plagwitz (Fds.), *Technology and the language teacher: The theoretical, historical, and practical interface* (pp. 248-261). Boston: Heinle & Heinle.
Liang, M.-Y. and C. J. Bonk (2009). Interaction in blended EFL learning principles and practice. *International Journal of Instructional Technology and Distance Learning, 6*(1): 3-15.
Littleton, K. & Häkkinen, P. (1999). Learning Together: Understanding the processes of Computer-based Collaborative Learning. In P. Dillenbourg (Ed.) Collaborative learning: cognitive and computational approaches (pp. 1-20). Pergamon: Oxford.
Liu, Y. B., & Zhao, S. H. (2008). Chinese language instruction in Singapore primary school classrooms: A comparative study. *Pedagogies: An International Journal, 3*, 168-186.
Looi, C. K. & Chen, W. (2010). Community-based individual knowledge construction in the classroom: A process-oriented account. *Journal of Computer Assisted Learning, 26*(3), 202-213.
Oxford, R. L. (1997). Cooperative learning, collaborative learning, and interaction: Three communicative strands in the language classroom. *The Modern Language Journal, 81*(iv), 443-458.
Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *The Journal of the Learning Sciences, 13*(3), 423-451.
Quintana, C., Eng, J., Carra, A.,Wu, H.-K., & Soloway, E. (1999). Symphony: A case study in extending learner-centered design through process space analysis. In M. G. Williams, M. W. Altom, K. Ehrlich, &W. Newman (Eds.), *Proceedings of CHI 99 Conference on Human Factors in Computing Systems* (pp. 473–480). Reading, MA: Addison-Wesley.
Reiser, B. J., Tabak, I., Sandoval,W. A., Smith, B. K., Steinmuller, F., & Leone, A. J. (2001). BGuILE: Strategic and conceptual scaffolds for scientific inquiry in biology classrooms. In S. M. Carver & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress* (pp. 263–305). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences, 13*(3), 273-304.
Sfard, A. (1998). On two metaphors for learning on the danger of choosing just one. *Educational Research, 27* (2), 4-13.
Shen, H. H. (2002). Level of cognitive processing: Effects on character learning among non-native learners of Chinese as a foreign language. *Language and Education, 18*, 167-182.
Shen, H. H. (2004). Level of cognitive processing: Effects on character learning among non-native learner of Chinese as a foreign language. *Language and Education, 18*, 17.
Sim, S.H. (2005). Teaching Chinese composition in Singapore secondary school. In: Shum, M.S. & Zhang, D.L. (Eds.), *Teaching Writing in Chinese Speaking Areas* (pp.245-258), Springer Netherlands.
Smith, B. (2004). Computer-mediated negotiated interaction and lexical acquisition. *Studies in Second Language Acquisition, 26*(3), 365-399.
Tabak, I. (2004). Synergy: a complement to emerging patterns of distributed scaffolding. *Journal of the Learning Sciences, 13*(3), 305-335.
Tan, C. (2006). Change and continuity: Chinese language policy in Singapore. *Language Policy, 5*(1), 41-62.
Tse. S.K. (2002). *Comprehensive and effective teaching and learning of Chinese characters*. Hong Kong: Greenfield Enterprise Ltd.
Vahey, P., Tatar, D., & Roschelle, J. (2007). Using handled technology to move between private and public interactions in the classroom. In M. van't Hooft & K. Swan (Eds.). Ubiquitous computing in education: Invisible technology, visible impact (pp.187-210). Mahway, NJ: Lawrence Erlbaum Associates.
Warschauer, M. (1997). Computer-mediated collaborative learning: Theory and practice. *The Modern Language Journal, 81* (4), pp. 470-481.
Warschauer, M., & Kern, R. (2000). *Network-Based Language Teaching: Concept and Practice*. Cambridge University Press.
Weinberger, A., Stegmann, K., & Fischer, F. (2007). *Knowledge convergence in collaborative learning: Concepts and assessment. Learning and Instruction, 17*, 416-426.
Wen. Y., Looi, C. K., & Chen, W. (2011). Who are the beneficiaries when CSCL enters into Second Language classroom. *Global Chinese Journal of Computers in Education,* 7. 91-107.
Wing, W. K., Ho, C. L., Chung, A. L. S., & Tse, S. K. (2003). Structural awareness, variation theory and ICT support. *Educational Studies in Language and literature, 3*, 53-78.
Wood, D., Brunet; J., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 17,* 89-100.

Acknowledgments

This material is based on the work supported by the National Research Foundation Singapore under Grant NRF2007-IDM003-MOE-001 and NRF2009-IDM001-MOE-019. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Research Foundation Singapore. We are grateful to the School of Science and Technology, and Mayflower Primary School for collaborating with us on this research.

Orchestrating Web-Based Collaborative Inquiry Learning with Small Group and Classroom Scripts

Ingo Kollar, Christof Wecker, Sybille Langer, Frank Fischer, Ludwig-Maximilians-University of Munich, Leopoldstrasse 13, D-80802 Munich
Email: ingo.kollar@psy.lmu.de, christof.wecker@psy.lmu.de, sybille.langer@psy.lmu.de, frank.fischer@psy.lmu.de

Abstract: Collaborative inquiry learning is a promising approach to foster students' online search competence. Yet, to be effective, it needs to be structured appropriately. In a quasi-experimental field study employing a 2x2 design, we investigated the effects of a small-group collaboration script (present vs. absent), two variants of a classroom script (online search activities constantly located on the small group level vs. online search activities alternating between plenary and small group level) and their different combinations on online search competence during an inquiry learning unit on Genetic Engineering. Results indicate that scaffolding on at least one of the two levels (realizing plenary phases or providing a small group script) is necessary to reach higher levels of online search competence. However, combining classroom scripts that alternate between plenary and small-group phases and small-group scripts did not further improve online search competence.

Introduction

Scientific literacy is the capacity to deal with issues in everyday life that are related to science (Laugksch, 2000). As science progresses continuously, no body of domain-specific knowledge acquired in school will be sufficient for this purpose throughout a lifetime. However, the Internet brings up-to-date scientific findings in the reach of everyone. Accordingly, scientific literacy has to include online search competence as an integral part. Yet, as prior research has shown, searching and finding relevant, credible and scientifically substantiated information on the Internet is a challenging task for high-school students (Lazonder, 2005). Thus, it is a pressing question how online search competence can be facilitated during secondary education.

One promising approach to achieve this is inquiry learning (e.g., de Jong, 2006), during which students act similarly to scientists who are confronted with an authentic research problem. However, prior research has demonstrated that inquiry learning needs to be structured to lead to significant learning (de Jong & van Joolingen, 1998), especially when it involves small-group collaboration, since students rarely collaborate on a high level spontaneously (Cohen, 1994). We argue that when collaborative inquiry learning is embedded in a real classroom situation, this guidance can be provided in at least two ways. First, if large portions of classroom learning are realized in small groups, it can effectively be supported with *small-group collaboration scripts* (e.g., de Wever, van Keer, Schellens & Valcke, 2010; Kollar, Fischer & Slotta, 2007; Weinberger, Ertl, Fischer & Mandl, 2005). Second, so-called *classroom scripts* that distribute learning activities over the different social planes of the classroom (e.g., the plenary, a small group and an individual level; see Dillenbourg & Jermann, 2007) may foster the acquisition of online search competence .In this paper we investigate the effects of different small group collaboration scripts and classroom scripts as well as their different combinations on the acquisition of online search competence in a 4.5 weeks inquiry learning unit on Genetic Engineering.

Collaborative Inquiry Learning as an Approach to Foster Online Search Competence

The basic idea of inquiry learning is that students should acquire domain knowledge and research skills by being confronted with science problems and attempting to solve them by engaging in scientific activities such as hypothesis generation, experimentation and drawing conclusions (de Jong, 2006). The retrieval of information in scientific publications or other sources and the use of this information for the construction of scientifically valid arguments may be regarded as yet another typical inquiry activity. Broken down to student learning and given the fact that today the Internet is the main source for information on science-related debates for high school students, having them search the Internet for such information to create arguments they can use in scientific debates can be seen as a kind of inquiry learning that is especially needed today.

In empirical research, inquiry learning has often been employed in a collaborative mode (i.e. involving small groups of learners), since engaging in scientific activities individually may be and often is too a demanding task. That said and considering results from prior research on collaborative learning indicating that learners rarely collaborate on a high level spontaneously (e.g., Cohen, 1994), there is a clear need for instructional guidance in order to make collaborative inquiry learning a successful experience. Accordingly, research on collaborative inquiry learning has intensively studied and identified ways to scaffold and support small groups during their inquiry (e.g., Sandoval, 2003; van Joolingen et al., 2005).

Scripts as Scaffolds for Collaborative Inquiry Learning

Based on empirical evidence that demonstrates that students often have difficulty to collaborate on a high level (e.g. because they lack collaborative competences; e.g. Cohen, 1994), CSCL research has over the last years rediscovered scripts as promising tools to provide such support. When reflecting about how to foster the acquisition of online search competence in real classrooms that consist of a teacher and 20 or more students, scripts target at least two levels: First, so called "small group collaboration scripts" (Kollar et al., 2006) specify, distribute and sequence learning activities and roles among the partners of a *small group* (e.g., dyads or triads). Second, the script idea may be expanded to the classroom level by introducing "classroom scripts" (very similar to Dillenbourg and Jermann's, 2007, "macro-scripts") as an instructional means to specify, distribute and sequence learning activities *over the different social planes of the classroom* (the plenary, the small group and the individual level). Both notions of scripts are described next.

Small-group Collaboration Scripts

CSCL research has developed and empirically investigated an impressive collection of small group collaboration scripts designed to raise the quality of *collaboration processes* (Hämälainen, & Arvaja, 2009; Rummel & Spada, 2005; Schellens, de Wever, van Keer & Valcke, 2005) and *individual learning outcomes* (Ertl, Kopp & Mandl, 2007; Kollar et al., 2007; Wecker & Fischer, 2007; Weinberger et al., 2005) of learners collaborating in small groups (e.g., dyads or triads). For example, such a script may give one learner of a dyad the task to suggest key terms for a collaborative online search and his/her partner the task to comment on the adequacy of these terms. In a next step the script may ask one learner to suggest a link to click on, while the other learner may be asked to estimate whether the proposed linked is likely to contain the needed information etc.. While there are some studies that investigated the effects of collaboration scripts on *collaboration processes and outcomes* in field settings (e.g., Hämälainen & Arvaja, 2009; de Wever et al., 2010), the effects of collaboration scripts on *individual learning outcomes* have up to now largely been investigated in laboratory studies employing rather short learning phases. For example, Weinberger et al. (2005) report on a study in which triads of learners collaborated in an asynchronous discussion board with the task to collaboratively analyze authentic problem cases by aid of a psychological theory. To support collaboration, they used a small-group collaboration script which distributed the roles of a case analyst and two criticizers among the learning partners. These roles rotated among the learners and each role was supported with appropriate prompts. Results indicated that compared to unstructured CSCL, the small-group collaboration script helped students acquire higher levels of domain-specific knowledge. Other studies demonstrated positive effects of small-group collaboration scripts on the acquisition of more domain-general competences. For example, Kollar et al. (2007) could show that learners who were supported by an argumentation-related small-group collaboration script in an inquiry learning environment reached higher levels of argumentation competence (as a learning outcome) than learners who were not supported by a small group script. However, since the reported studies were conducted in lab contexts, it is unclear whether small group collaboration scripts will also lead to positive individual learning outcomes when used in field settings employing longer learning phases.

Classroom Scripts

When thinking about how to instructionally support learners during inquiry learning under real classroom conditions, i.e. in classrooms that involve a teacher and 20 or more students, scaffolding of small groups is only one of several options. In an expansion of the script concept to the whole classroom (as compared to the small group level only; see Dillenbourg & Jermann, 2007), activities and roles may also be specified, sequenced and distributed *over social planes of the classroom beyond the small group level*, namely also over the plenary and the individual level. For example, a lesson may start with the teacher modelling how to search for evidence on the Internet in front of the class (plenary activity), followed by student dyads searching for evidence together (small-group activity) and closing with a final phase in which students search the Internet individually (individual activity). In the following, such instructional interventions that structure learning activities through their distribution of different social levels will be called *classroom scripts*. It should be noted, however, that the term "classroom scripts" has in the past also been used in a descriptive meaning for culturally shared norms about interaction patterns occurring in a classroom that emerge without explicit external interventions (see Seidel et al., 2002). In contrast to this, we use the term in a prescriptive sense to refer to explicit instructional interventions that change the structure of activities in larger social entities like classrooms in the way described.

When taking a look at prominent instructional approaches from the literature, it is striking that very often a distribution of activities over the different social planes of the classroom is proposed (although not always over all three levels), i.e. research already provides a collection of classroom scripts designed for different purposes. For example, during Problem-based Learning (PBL; see for example Hmelo-Silver, 2004), small student groups and a tutor meet on a plenary level to jointly discuss an authentic problem case (e.g., a description of a patient with particular illness symptoms), and then split up to individually acquire knowledge on physiological processes that may account for the illness. After that, students come back to groups to discuss what they have found during their individual study and apply this knowledge to the case. Similar activity structures can be found in Reciprocal Teaching (Palincsar & Brown, 1984) and Learning by Design (Kolodner, 2007) Although the distribution of activities over different social planes obviously is regarded important for learning in these approaches, it is striking that there is hardly any research that systematically compares the effects of different classroom scripts differing in the way activities are distributed over the different social planes on learning processes and outcomes while other features of the classroom script are held constant. Therefore, it is an open empirical question how to best sequence and distribute learning activities during inquiry learning over the different social planes in order to help students reach high levels of competence.

Combining Small Group and Classroom Scripts - Synergistic Scaffolding or Over-scripting?

Based on theoretical considerations, specific combinations of small group and classroom scripts may be regarded as either beneficial or detrimental for learning. The expectation that a combination of a particular small-group collaboration script and a particular type of classroom script will lead to better learning outcomes than any of the two alone would be an example for “synergistic scaffolding” (Tabak, 2004). Synergistic scaffolding is realized when two different scaffolds mutually amplify their each specific effects on the same learning outcome. Applied to the notion of small group and classroom scripts, synergistic scaffolding would be realized when a combination of a small group script and one specific variant of a classroom script leads to higher competence levels than any of the two scripts alone.

However, two scaffolds may also inhibit each other’s effects on competence acquisition and thus lead to what Dillenbourg (2002) called “over-scripting”. In particular, it may be that adding a small group script to a classroom script that also includes modeling by the teacher (as a plenary activity (e.g., Rummel & Spada, 2005) may put too many constraints on the learners so that productive and creative search processes are undermined. This would result in less positive learning outcomes than if either a small-group collaboration script or a classroom script including plenary phases would have been presented alone.

Since prior research on the effects of different combinations of small group collaboration scripts and classroom scripts in the context of CSCL is scarce, it is hardly possible to opt for one and reject the other expectation right away. Therefore, in our empirical study we investigated not only the separate effects of a small group collaboration script and different variants of a classroom script, but also their combinations.

Research Questions and Hypotheses

We investigated the effects of a small-group collaboration script (present vs. not present) and two variants of a classroom script as well as their different combinations on the acquisition of online search competence during a 4.5 weeks inquiry learning curriculum unit. In one variant, the classroom script located online search activities solely on the small group (i.e., dyadic) level (“Small-group Level Only-“ or “SLO classroom script”). The SLO classroom script can be regarded as the standard classroom script that would be employed when dyadic inquiry learning is realized. This was compared to a classroom script that realized online search activities alternating between the small group and the plenary level (“Alternations between Plenary and Small group level-“ or “APS classroom script”). Online search processes on a plenary level were realized as phases in which online search was either modelled by the teacher and a student or two students in front of class. Our research questions were:

1. How do small group (present vs. absent) and classroom scripts (SLO vs. APS) play together to support online search competence in a Web-based collaborative inquiry environment?
2. When no small-group collaboration script is provided, what are the effects of an APS classroom script compared to the effects of an SLO classroom script on online search competence in a Web-based collaborative inquiry environment?
3. When the employed classroom script locates all search activities on the small group level (SLO classroom script), what are the effects of a small group collaboration script compared to non-scripted dyadic online search in a Web-based collaborative inquiry environment?

Concerning research question 1, we could not establish a clear hypothesis due to the lack of prior similar research. Two conflicting hypotheses can be justified: On the one hand, the combination of the APS classroom script and a small-group collaboration script may produce “synergistic scaffolding” (Tabak, 2004) and thus have the most positive effect on the acquisition of online search competence. On the other hand, this combination may produce “over-scripting” (Dillenbourg, 2002), i.e. too many constraints for the learners to perform high-level online searches and thus lead to lower levels of online search competence than each of the two scripts alone. With respect to research question 2, we hypothesized that when no a small-group collaboration script is provided, learners working with the APS classroom script acquire higher levels of online search competence than learners working with the SLO classroom script. Concerning research question 3, we expected that under standard classroom script conditions (i.e., the SLO classroom script),learners collaborating on the basis of a small group script reach higher levels of online search competence than students not supported by a small group collaboration script.

Method

Sample and Design

174 9th graders from eight secondary urban school classrooms in Germany participated in a quasi-experimental field study. 90 students were female, 84 students were male. We implemented a 2x2-factorial pre-post test design with the independent factors “type of classroom script” (SLO vs. APS) and “small-group collaboration script” (present vs. absent; see table 1). The comparably small N in the condition “without small-group collaboration script/APS classroom script” was due to a higher drop out due to illness and other factors beyond our influence in the two classes in which this treatment was realized than in the other classes. In fact, all subjects who missed one or more lessons over the 4.5 weeks curriculum unit were excluded from data analysis. Originally, the numbers of students in the aforementioned condition was 46 and thus comparable to the Ns in the other three conditions.

Table 1: Design of the empirical study.

		Small-group collaboration script	
		Without	With
Type of classroom script	Search activities solely located on the small group level (SLO)	N = 43 students (2 classes)	N = 52 students (2 classes)
	Search activities alternating between small group and plenary level (APS)	N = 28 students (2 classes)	N = 51 students (2 classes)

Instructional Setting

For the purpose of the study, a 4.5-weeks inquiry-based curriculum unit on Genetic Engineering was developed. The students’ task was to develop a scientifically sound and valid position on whether “green” Genetic Engineering (i.e., the genetic modification of plants for food production) should be allowed or not. For this purpose, each student was equipped with a laptop computer which was connected to the Internet via a wireless LAN network. Overall, the unit spanned seven Biology lessons, which were led by the regular Biology teacher of the class, plus one pre- and one post-test lesson in which learners’ online search competence (pre and post instruction) as well as demographic variables were assessed. The actual unit started with an introduction to Genetics and Genetic Engineering. Also, the teachers in all classes introduced the students to how to perform a successful online search. The core of the seven lessons were three consecutive content-specific cycles on different aspects of Genetic Engineering (one on economic, one on ecological, and one on health aspects of Genetic Engineering). Each of these content-specific cycles included three steps. First, student dyads were asked to browse through a web-based project library which included biological knowledge on Genetics and Genetic Engineering. This online library was implemented within the Web-based Inquiry Science Environment (WISE; Slotta & Linn, 2009) and was designed on the basis of regular 9th grade Biology text books. In step 2, either student dyads or (in case of the APS classroom script) the teacher or the student modelers collaboratively formulated an initial argument (e.g., that eating genetically modified food is dangerous for health) and searched the Internet for evidence that would support, modify or discard their initial argument. Online search activities were supported by a software named S-COL (Wecker et al., 2010) which allowed for collaborative Internet browsing, i.e. that during their online search, both learning partners of a dyad always saw the same web sites, no matter who of the two clicked on a link or entered search terms. In step 3, the teacher led a plenary discussion in which students brought together their arguments they were supposed to back up with their Internet findings. To reduce statistical dependencies in the sample, dyads were re-organized before each new search phase.

Independent Variables

Both script variables were during each of the three online search phases. The basis of the two treatments (the small-group and the classroom script) was a multiple-step online search process that was either to be performed by all student dyads or (during modeling phases) by the teacher together with one student or by one teacher-selected student dyad in front of the whole class. The search process to be adopted by the students was derived from a comprehensive literature review, a cognitive task analysis and an expert-novice comparison of Internet search activities using think-aloud methodologies prior to this study (see Kollar, Wecker & Fischer, 2009). Basically, both treatments split the dyadic online search processes up into a five-step strategy: (1) the *formulation of an initial argument and a sketch of the information needed*, (2) the *selection of search terms*, (3) the *evaluation of the hit list*, (4) the *localization of relevant information* on a web page, and (5) a *written formulation of the final elaborated argument*. For example, during the second step, the *selection of search terms*, learner A was prompted to suggest a set of search terms, while B was asked to first recall the information they had decided to look for, and comment on the adequacy of A’s suggestions for the search terms.

In the *APS classroom script*, online search phases consisted of both dyadic online searches, during which all dyads performed a collaborative online search, and modeling phases in which the teacher picked either a student to model a successful online search with in front of the whole class, thereby employing specific steps of the search strategy just described, or pulled one student dyad out of the group to perform an online search in the plenary, while commenting on the quality of their online search process. That way, an alternation between search activities on the small-group and the plenary level was realized. The moments for each modeling phase were roughly chosen to be close to when all student dyads were performing the search steps that were modeled: Modeling of the first two search steps happened at the beginning of the first search activity, modeling of the third and fourth step happened at about half time of the second search activity, and modeling of the final step happened closer to the end of the third search activity. In the *SLO classroom script*, no such modeling phases were employed. Rather, students used the complete time allocated for online search within dyads.

The *small-group collaboration script* was technically implemented with the S-COL software (Wecker et al., 2010) and distributed cognitive and metacognitive activities related to each single step in the search process among the members of each dyad. To do so, S-COL divided the browser window in two frames (see Fig. 1). In the right part of the screen (the browsing area), students could view regular web. In the left part of the screen (the scaffolding area), each single learner received particular script prompts that guided him or her in what exactly to do during this step of search. To guarantee the display of "correct" prompts for each step, a software algorithm was used that was able to differentiate between (a) the Google start page, (b) the Google hit list or (c) any other web page. For example, when S-COL recognized the Google start page (i.e. during the selection of search terms), learner A receivec prompts such as "Please suggest a couple of search terms to your learning partner", while learner B was prompted "Listen to your partner's suggestions and estimate how well his/her search terms are suited to find what you are looking for". The two roles were switched after each search cycle. In the condition without small-group collaboration script, dyads performed unstructured online searches, i.e. their browsers were connected, but no scaffolding area appeared on the screen.

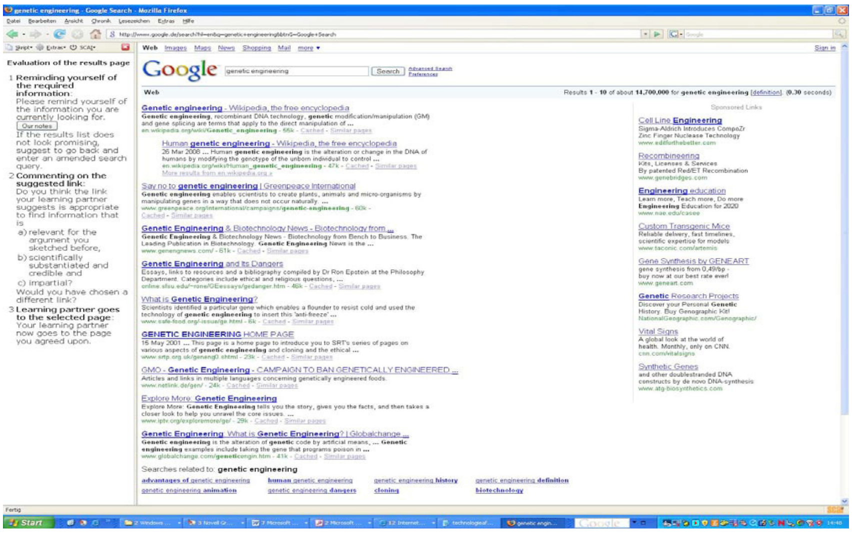


Figure 1. Screenshot of the Small-group Collaboration Script (left frame: prompts for one of the two learners; the learning partner at the same time received complementary prompts; right frame: regular Google hit list).

Dependent Variable

Online search competence was measured by an individual test that asked students to describe in as much detail as possible how they would proceed if they had to use the Internet to form a position on a science topic that was different from Genetic Engineering. In the pre-test, the topic was whether mobile phone transmitters should be banned from the immediate surroundings of kindergartens and day nurseries; in the post-test, the topic was whether nuclear power plants should be shut down. The test pre-structured the students' answers by using a number of lines and two columns. In the first column of the first line, participants were supposed to write the action they would take first, and in the second column to note quality criteria they would apply while taking this step. This could be repeated up to eight times for the following steps. A general expert solution, based on the steps of the to-be-acquired online search strategy (see above), guided the coding of the students' answers for the occurrence of each individual step and quality criterion contained in the expert solution. Two coders independently rated 15 % of the material, reaching a sufficiently high inter-rater agreement (ICC = .83 for the post test, ICC = .51 for the pre test).

Statistical Analyses

The alpha-level for all analyses was set to 5 %. To answer research question 1 (combination of small-group and classroom scripts), an analysis of covariance with the two script variables as fixed factors, classes as fixed factor nested within experimental conditions, and the scores in the online search competence post test as dependent variable were conducted. Pre test scores were used as a covariate. With respect to hypotheses 2 and 3, planned contrasts were conducted to compare the effects of the APS vs. SLO classroom scripts in the condition without small group collaboration script (hypothesis 2) and the effects of the conditions with vs. without small group collaboration script when the SLO classroom script was used (hypothesis 3). To avoid inflation of Alpha errors, Bonferroni corrections were applied, i.e. the two planned contrasts were tested on an alpha-level of 2.5 %.

Results

Table 3 shows the mean post test scores and standard deviations in the online search competence tests (note that experimental conditions did not differ significantly with respect to online search competence in the pre-test; $F(1,166) < 2.20$; *n.s.*). Descriptively, students who received both modeling of search strategies and had the opportunity to run online searches on their own (APS classroom script condition), but without the small-group collaboration script reached the highest post test scores. At the lower end of the spectrum were learners whose search activities were constantly located on the dyadic level (SLO classroom script condition) and who did not receive a small-group collaboration script. The means of the other two groups were about half-way in between.

Table 3: Means and standard deviations concerning online search competence in the post test in the four experimental conditions.

	Without small group collaboration script				With small group collaboration script			
	SLO classroom script		APS classroom script		SLO classroom script		APS classroom script	
	M	*SD*	M	*SD*	M	*SD*	M	*SD*
Online search competence (post test)	2.65	*2.27*	4.75	*2.15*	3.90	*2.58*	3.96	*2.54*

To the test the two competing hypotheses related to research question 1 (synergistic scaffolding hypothesis vs. over-scripting hypothesis), an ANCOVA revealed a significant interaction effect ($F(1,165) = 12.41$; $p < .01$; partial $Eta^2 = .07$), indicating that the small-group collaboration script was more helpful when the classroom script located all search activities on the small group level (SLO classroom script) than when search activities were realized alternately on the plenary and the small group level. When the classroom script alternated search activities between the small-group and the plenary level (APS classroom script),the small-group collaboration script seemed to have no effect. Based on this significant interaction effect, main effects of the two treatments were not further examined.

To test *hypothesis 2* ("If not provided with a small-group collaboration script, learners working with the APS classroom script acquire higher levels of online search competence than learners working with the SLO classroom script"), a planned comparison revealed a significant difference between the APS and the SLO classroom script when no small-group collaboration script was provided ($F(1,61) = 23.05$, $p < .01$, partial $Eta^2 = .27$). Thus, hypothesis 2 was confirmed. To test *hypothesis 3* ("If the SLO classroom script is employed, learners collaborating on the basis of the small group collaboration script reach higher levels of online search competence than students not supported by a small group collaboration script"), a planned comparison revealed a significant difference between the conditions with vs. without small group collaboration script when the SLO classroom script was employed ($F(1,88) = 8.89$, $p < .01$, partial $Eta^2 = .14$). Thus, hypothesis 3 was supported.

Discussion

In this study, we investigated to what extent a small-group collaboration script and particular types of classroom scripts as well as their different combinations facilitate the acquisition of online search competence in the context of a multiple-weeks web-based collaborative inquiry curriculum unit on Genetic Engineering. With respect to our first research question, namely how a small-group collaboration script and a classroom script that alternates between plenary and small-group activities (APS classroom script) play together to support the acquisition of online search competence, we established two competing hypotheses. On the one hand, we argued that combining the small-group script with the APS classroom script might represent an example of "synergistic scaffolding" (Tabak, 2004) and lead to the highest results in terms of students' acquisition of online search competence. On the other hand, we argued that this combination may produce "over-scripting" (Dillenbourg, 2002) such that this combination would lead to lower effects than each of the two treatments alone.

Upon close inspection of the results, both hypotheses need to be rejected: Clearly, no synergistic scaffolding effect was found because in the presence of the APS classroom script students working with the small group collaboration script did not reach higher scores in the posttest than learners who were not provided with a small group collaboration script. However, also no over-scripting effect has been observed: Adding an APS classroom script to a small group collaboration script neither enhanced nor harmed the effects of the small group script. Thus, our results rather indicate a kind of "functional equivalence" of small-group and classroom scripts: As the results related to hypotheses 2 and 3 show, both treatments alone had comparably large positive effects on online search competence when the other treatment was not used. When used in combination, however none of the two treatments adds substantially to the other.

Obviously, alternating search activities between a small group and a plenary level is a better way of distributing search-related activities over the classroom than having students perform such activities solely on the small group level. This adds to earlier research that has demonstrated the value of modeling on student learning (e.g., Palincsar & Brown, 1984; Rummel & Spada, 2005). Likewise, if students are supposed to perform online search activities in dyads without intermitting modeling phases, then supporting them with a small-group collaboration script is necessary. This corroborates findings from prior studies that have demonstrated the effectiveness of small-group collaboration scripts on individual learning outcomes (Ertl et al., 2007; Weinberger et al., 2005). However, combining the two scaffolds neither made learners perform worse than students in the control condition (without small group collaboration script and the SLO classroom script), which would have been expected following the over-scripting hypothesis, nor made learners perform on a higher level than could have been expected by adding the two individual effects as the synergism hypothesis would have predicted.

With respect to the differential effects of the two employed classroom scripts, our study is among the first ones to demonstrate that specific distributions over the different social planes of the classroom may have differential effects on student learning. This assumption has already been reflected in the classroom layouts of prominent instructional approaches such as Problem-based Learning (Hmelo-Silver, 2004), Learning-by-Design (Kolodner, 2007) or Reciprocal Teaching (Palincsar & Brown, 1984), but has hardly been systematically investigated before. Obviously, interweaving plenary activities into small-group activities is a good strategy to bring the expertise of the teacher to bear in (a) modeling a to-be-acquired strategy or competence and (b) in correcting developing misconceptions or shallow strategies that are used and beginning to be internalized by the students. Yet, these interpretations may be confirmed by process analyses that demonstrate direct sequential effects of modeling or of discussing small group script prompts on subsequent search activities of the dyads. Such analyses are currently under way.

In interpreting the results of our study, two *limitations* should be noted. First, the online search competence test that was employed did not ask students to really perform an online search themselves. Rather, the test asked learners to describe what they would do when they had the task to search the Internet to develop a well-grounded position in a science debate. Possibly, having learners actually perform a new search on a different science topic would yield somewhat different results. This should be tested in future research. Second, it should be noted that our variation on the classroom script level only included two social levels, the plenary and the small-group level. However, since the competence that we wanted students to acquire is in the end an individual one, it may be promising to also include phases on an individual level. It is not unlikely that, to reach a high level of online search competence, online searches should also be practiced individually - possibly with further guidance on the individual level. Also this question is subject to future research.

Yet, our study produced promising results which may be of value for educational practice. For teachers, it is very often a hard task to structure small-group activities. Our results indicate that such efforts may be neglected when instead small-group collaboration is from time to time interrupted by plenary activities during which the teacher sets the small groups "on the right track" again. Compared to how to structure small-group collaboration, teachers are very often more experienced in how to design such plenary level phases. Thus, the results of our study should encourage teachers to use their skills in designing high-level plenary activities and align them with small-group learning phases, without structuring the latter ones too severely.

References

Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research, 64*(1), 1-35.

de Jong, T. (2006). Technological advances in inquiry learning. *Science, 312,* 532-533.

de Jong, T. & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research, 68*(2), 179-201.

De Wever, B., Van Keer, H., Schellens, T., & Valcke, M. (2010b). Structuring asynchronous discussion groups: Comparing scripting by assigning roles with regulation by cross-age peer tutors. *Learning and Instruction, 20,* 349-360.

Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. In P. A. Kirschner (Ed.), *Three worlds of CSCL. Can we support CSCL* (pp. 61-91). Heerlen: Open Universiteit Nederland.

Dillenbourg, P. & Jermann, P. (2007). Designing integrative scripts. In F. Fischer, I. Kollar, H. Mandl &, J. Haake (Eds.), *Scripting computer-supported communication of knowledge. Cognitive, computational, and educational perspectives* (pp. 259-288). New York: Springer.

Ertl, B., Kopp, B., & Mandl, H. (2007). Supporting collaborative learning in videoconferencing using collaboration scripts and content schemes. In F. Fischer, H. Mandl, J. M. Haake, & I. Kollar (Eds.), *Scripting computer-supported communication of knowledge – cognitive, computational, and educational perspectives* (pp. 213-236). Berlin: Springer.

Hämäläinen, R. & Arvaja, M. (2009). Scripted collaboration and group-based variations in a higher education CSCL context. *Scandinavian Journal of Educational Research, 53*(1), 1-16.

Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review, 16,* 235-266.

Kollar, I., Fischer, F., & Hesse, F. W. (2006). Collaboration scripts - a conceptual analysis. *Educational Psychology Review, 18*(2), 159-185.

Kollar, I., Fischer, F., & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning & Instruction, 17* (6), 708-721.

Kollar, I., Wecker, C., & Fischer, F. (2009, August). *Supporting high-school students' scientific literacy through scripted collaborative inquiry - a use-inspired basic research program*. Paper presented at the 13th Biennial Conference for Learning and Instruction der European Association for Research on Learning and Instruction (Earli) „Fostering Communities of Learners", August 25 to 29, 2009.

Kolodner, J. L. (2007). The roles of scripts in promoting collaborative discourse in learning by design. In F. Fischer, I. Kollar, H. Mandl & J. M. Haake (Eds.), Scripting Computer-Supported Collaborative Learning - cognitive, computational, and educational perspectives (pp. 237-262). New York: Springer.

Laugksch, R. C. (2000). Scientific Literacy: A conceptual overview. *Science Education, 84,* 71-94.

Lazonder, A.W. (2005). Do two heads search better than one? Effects of student collaboration on Web search behavior and search outcomes. *British Journal of Educational Technology, 36*, 465-475.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1*(2), 117-175.

Rummel, N., & Spada, H. (2005). Learning to collaborate: An instructional approach to promoting collaborative problem-solving in computer-mediated settings. *The Journal of the Learning Sciences*, *14*(2), 201-241.

Sandoval, W. A. (2003). Conceptual and epistemic aspects of students' scientific explanations. *The Journal of the Learning Sciences, 21*(1), 5-51.

Schellens, T., Van Keer, H., Valcke, M., & De Wever, B (2005). *The impact of role assignment as scripting tool on knowledge construction in asynchronous discussion groups.* Paper presented at the International Conference on Computer Supported Collaborative Learning (CSCL). Taipei, Taiwan.

Seidel, T., Prenzel, M., Duit, R., Euler, M., Geiser, H., Hoffmann, L., Lehrke, M., Müller, C. T. & Rimmele, R. (2002). „Jetzt bitte alle nach vorne schauen!" - Lehr-Lernskripts im Physikunterricht und damit verbundene Bedingungen für individuelle Lernprozesse. *Unterrichtswissenschaft, 30*(1), 52-77.

Slotta, J. D. & Linn, M. C. (2009). *WISE Science: Inquiry and the Internet in the Science Classroom.* Teachers College Press.

Tabak, I. (2004). Synergy: A complement to emerging patterns of distributed scaffolding. *The Journal of the Learning Sciences, 13*(3), 305-335.

Van Joolingen, W. R., de Jong, T., Lazonder, A. W., Savelsbergh, E. R. & Manlove, S. (2005). Co-Lab: research and development of an online learning environment for collaborative scientific discovery learning. *Computers in Human Behavior, 21,* 671-688.

Wecker, C. & Fischer, F. (2007). Fading scripts in computer-supported collaborative learning: The role of distributed monitoring. In C. Chinn, G. Erkens & S. Puntambekar (Eds.), Proceedings of the CSCL 2007. Rutgers, The State University of New Jersey, New Brunswick, New Jersey, USA, July 16th – July 21st, 2007 (pp. 763-771). International Society of the Learning Sciences.

Wecker, C., Stegmann, K., Bernstein, F., Huber, M.J., Kalus, G., Kollar, I., Rathmayer, S., & Fischer, F. (2010) S-COL: A Copernican turn for the development of flexiblyreusable collaboration scripts. *International Journal of Computer-Supported Collaborative Learning,* 5(3), 321-344.

Weinberger, A., Ertl, B., Fischer, F., & Mandl, H. (2005). Epistemic and social scripts in computer-supported collaborative learning. *Instructional Science, 33*(1), 1-30.

Acknowledgments

The research reported here was funded by the Deutsche Forschungsgemeinschaft (DFG). The authors would like to thank the students and teachers who have participated in the study.

Scripting Collaborative Learning in Smart Classrooms: Towards Building Knowledge Communities

Michelle Lui, Mike Tissenbaum, James D. Slotta, University of Toronto, 252 Bloor Street West Toronto, Canada, ON,
Email: michelle.lui@utoronto.ca, mike.tissenbaum@utoronto.ca, jslotta@oise.utoronto.ca

Abstract: This paper shares preliminary findings on a new program of research on collaborative learning in smart classrooms. Using a co-design method, researchers worked with high school teachers to create engaging curriculum activities that provided the context for two studies in math and physics. The activity designs aim to increase the depth of students' conceptual understanding by breaking down learning goals into manageable sections. Students "tagged" questions in terms of relevant concepts, analyzed visualizations that captured the collective wisdom of the classroom community, critiqued results, and negotiated a shared understanding of domain-specific principles. Twenty-one mathematics students from grades ten and eleven participated in the first study; thirty-two grade twelve physics students participated in the second. Results showed improvements in problem-solving (in the second study), as well as improved tagging proximity to an expert model (in both studies). Issues with collaboration scripts used in the smart classroom are also discussed.

Introduction

Students' lives are increasingly being shaped by technology, and their future in the 21st century workplace will surely demand a deep fluency with information technologies, the most exciting of which are only just emerging (e.g., ubiquitous computing, physical computing, real-time collaboration environments). Yet the primary model of instruction in today's classrooms remains one of traditional didactic pedagogy – particularly in math and science – where lectures, problem sets and exams rule the day, and peer competition is more likely than collaboration. Designing instruction to promote deep understanding means adopting new modes of inquiry where students are encouraged to think deeply about materials and develop their own understandings (Linn & Eylon, 2006; Quintana et al., 2004). Technology environments can help scaffold students and teachers in new forms of learning where they collaboratively engage with materials, coordinated by scripts that help guide the flow of people, materials, and activities (Kollar, Fischer, & Slotta, 2007). However, current means of supporting instruction with technology are typically isolating, with students hidden behind rows of large monitors, as seen in typical computer labs.

The idea of forming knowledge communities, where members are given the responsibility to generate and build upon each other's ideas, ultimately developing their own knowledge base, is becoming more relevant for today's learners who are tasked with establishing collaboration skills, critical thinking and communication skills, in addition to learning content knowledge (Bereiter & Scardamalia, 2003; Partnership for 21st Century Skills, 2009). With the advent of sophisticated collaboration technologies, it is now possible to develop applications that allow for more seamless and dynamic collaboration, supporting real-time face-to-face interactions while maintaining an intelligent knowledge base that represents the collective wisdom of the entire class. This paper presents an ambitious new program of research that investigates how technology-enhanced learning environments can be embedded deeply within the classroom: becoming more visible and yet less intrusive, responding intelligently to student inputs, and capturing the collective wisdom of the classroom community as a resource for all participants. We have developed a new "smart classroom" and applied it in a design partnership with high school mathematics and physics teachers in order to address students' deep conceptual understanding of these traditionally challenging domains. Sections below describe our curriculum designs in two smart classroom activities, our technology and material designs, and the outcome of our studies.

Smart Classrooms

Our research recognizes the potential of technology-enhanced learning environments to enable new forms of learning, where students collaborate within their classroom or across multiple classrooms, dynamically generating knowledge, building on peer ideas, and investigating questions as a knowledge community. Our notion of a smart classroom employs a wide range of technologies to allow investigations of a full spectrum of collaborative inquiry and knowledge construction activities. Our technology framework consists of a portal that allows students to register and log in, an intelligent agent framework that allows tracking of student interactions, a central database that houses curriculum materials and the products of student interactions, and a visualization layer that controls the information presented to students (Slotta, 2010). At present, our implementation of a smart classroom includes four large projected displays in each corner of a classroom, a fifth, larger multi-touch display on the front wall, and twenty laptops, all interconnected via high-speed wireless network.

Knowledge Communities

Knowledge communities are created through collaboration, where students learn through generating and building upon each other's ideas within their unique knowledge base (Bereiter & Scardamalia, 2003). Brown and Campione (1990; 1996) first offered an interpretation of science classrooms as knowledge communities in their Fostering Communities of Learners research program, where students engaged in research activities, shared findings, and used their collective expertise to develop a shared understanding. A related approach called Knowledge Building instead values advancement or creative work performed as a community (Scardamalia & Bereiter, 2006). In this perspective, students are not recognized for what is in their minds but for contributions they make to the group's knowledge base. Students are more explicitly involved in the process of building upon one another's ideas, with the goal of developing a knowledge community in the classroom (Scardamalia and Bereiter, 1994; 1996). We envision participation in the smart classroom as a portal to a collective knoweledge base. Students engaging in collaborative activities within the smart classroom can use this knowledge base to step back and reflect on their participation, and allow their shared knowlege to emerge over time. Unfortunately, there are numerous challenges in implementing a knowledge community approach, which includes the elevated investment required of teachers to use it, as well as scalability issues for replication in research (Slotta & Peters, 2008). The open-ended nature of topics covered is also unsuitable to secondary school science courses, where teachers often carry heavy curriculum expectations (Peters & Slotta, 2009).

Orchestration of Complex Collaborative Designs

Prior research on supporting complex collaborative learning designs offers some insight into the affordances of new technologies (e.g., awareness or data mining). One common approach is that of the "collaboration script," which has been shown to effectively foster collaboration and improve learning outcomes (e.g., De Wever, et al., 2009; Kollar, Fischer, & Slotta, 2007; Rummel & Spada, 2005; Weinberger, et al. 2005). The script serves to define participants, activities, roles, and groups, specifying how tasks are distributed, how groups are formed and the sequence of activity flow (Kobbe, et al., 2007). Collaboration scripts add structure to activities and distribute responsibility, ensuring that all members actively participate. Scaffolds designed to complement the script can help guide students' cognitive processes (O'Donnell, 1999), as well as their interactions with peers (i.e., from different roles within the script) by providing specific task instructions, modeling, and instruction on skills or methods (King, 1999). Scripts have been found to foster domain-general knowledge, such as argumentation (Stegmann, et al., 2007) and interdisciplinary communication (Rummel & Spada, 2007). Collaboration scripts offer smart classroom curriculum designs structure to maintain focus on specific roles and content expectations, while allowing room for students' free flowing ideas to emerge and be built upon.

Objective

To understand how certain aspects of a smart classroom and carefully designed complex collaboration activities might support learning in a knowledge community, we first investigated how students used collaborative discourse and collective knowledge to make connections among abstract domain concepts in math and physics. Through the use of visualizations as representations of shared knowledge, "tags" to connect concepts, and reflection scripts to enhance understanding, our research team created two iterations of an analogous pedagogical design in sequence, with findings from the first incorporated into the second design. This paper presents findings from both studies.

Method

Our research employs a design-based methodology, which is characterized by iterative cycles of design, evaluation and revision of an intervention for study in authentic settings (Brown, 1992; Design-Based Research Collective, 2003). Using a co-design method (Penuel, Roschelle, Shechtman, 2007), our team of researchers, designers, technology developers, and high school teachers met regulary to develop curriculum activities, content materials and specialized software to coordinate the flow of participants and media amongst various computers, servers, and displays. The studies took place in the aforementioned smart classroom, located at a private, local urban high school, where the excellent students, the creative and reflective teachers, and the supportive administration make this an ideal setting for an ongoing co-design partnership for educational innovations.

Study 1: Connecting Mathematical Concepts

The curriculum, detailed below, was designed to engage several small groups of students working in parallel as they "tagged" a common set of math problems. In so doing, a collaborative visualization emerged as the software synthesized tags from all groups. A set of thirty problems was developed by the teacher. Each problem may be classified into one or more of four category groups: Algebra & Polynomials, Functions & Relations, Trigonometry, and Graphing Functions. The basic goal of this activity was to help students understand the relationships between these four aspects of mathematics by having them visualize the association of math problems with multiple categories.

Participants

A total of twenty-one student volunteers in grades ten and eleven from our research partner high school participated in the study. Nineteen students participated in the pre-test and curriculum activities. Ten students completed the post-test, two of which did not participate in the curriculum activity. The curriculum activity was co-designed with a high school mathematics teacher.

Design

Part A - Tagging

Students were provided with individual laptops and logged in to a specially-designed system that coordinated the flow of activities for the session. The software agents distributed students into four groups, each specializing in one of the four mathematics categories. Each group gathered in a specified area of the room, where the individual students within that group were presented with a set of math problems. For each problem, the student decided whether it should belong in their group's category, without solving the problem. A projection screen in front of each group displayed a semantic map of that group's category, with each of the math problems as nodes. If a problem was tagged with their group's category, a connection between the problem and the category was made (Figure 1). This was represented on the screen as a line, linking the problem node to the category node. At the front of the room was a larger screen that showed the collaborative visualization, which displayed an aggregate of the connections made between the questions and all four categories.

Part B - Solving

Once the four groups reviewed all the questions, they were provided with pencil, paper, and calculators. The software agents presented students with only those problems that were tagged as belonging to their group's category as well as at least one other category. Students were then instructed to solve each problem, working as a group. Once completed, they took a photo of their solution using the laptop camera and uploaded it to the system.

Part C – Reflecting

Next, the students in each group were shown the tags that other groups had assigned to the questions they just solved and asked to vote on the credibility of those tags. For example, students in the "Algebra & Polynomials" group might be presented with a problem that had been tagged by their group as well as the "Trigonometry" group. Students would be informed that the problem had been tagged as being "Trigonometry" as well, and asked whether they agreed with the "Trigonometry" connection. After stating their agreement or disagreement, the students would then be prompted to explain their choice in reflection notes. The visualizations were updated in real-time – the lines representing connections that had stronger consensus became thicker, and uploaded solutions appeared as new nodes.

Procedure

Students were given a paper-based pre-test of twelve questions. Rather than solve the questions, they were asked to identify problems as "Algebra & Polynomials", "Functions & Relations", "Trigonometry", and/or "Graphing Functions". They could check off more than one category, and were also asked to point out any other themes the questions were related to. These were collected and compared with the same task as performed by the teacher, who had helped to assemble the problems. The post-test was similar to the pre-test, except it was in the form of an online survey and a new set of problems were used. Boxes for comments were made available for students' input on their experience on the pre-test, curriculum activity and post-test. Qualitative analysis was performed on video captured during the activity and audio data collected at group workstations.

Results

For the pre- and post-test, as well as evaluation of the tagging from part A of the curriculum design, we examined two constructs, accuracy and structuredness. Accuracy scores were compiled by looking at the group average of correct connections against the total number of connections made by the teacher. Following previous work using mathematics concept maps (Hasemann & Mansfield, 1995), the "structuredness" for each set of data was evaluated with the number of connections made compared against the total number of potential connections. On the pre-test, students achieved an accuracy rate of 72% (*SD* = 6.16); the structuredness level was 52% (*SD* = 15.70). During the curriculum activity, students made more connections between problems and categories, with an accuracy rate of 80% (*SD* = 17.41) and a structuredness rate of 80% (*SD* = 23.44). On the post-test, the accuracy rate for the smart classroom group was 77% (*SD* = 6.04), and the structuredness rate was also 63% (*SD* = 19.16). For those who did not participate in the activity, the accuracy was 76% (*SD* = 4.82), while the structuredness rate was 50% (*SD* = 5.89), as shown in Figure 2.

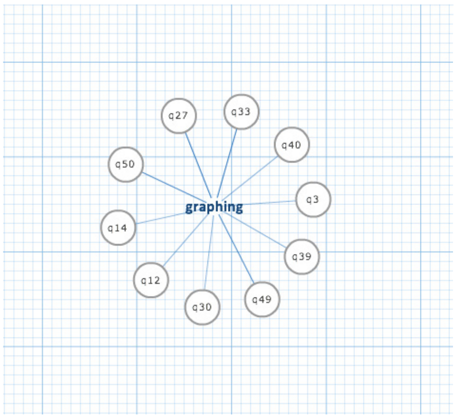


Figure 1. Tags Made by the Graphing Group.

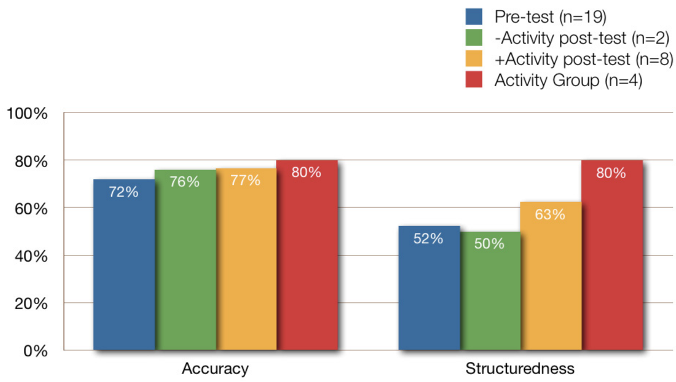


Figure 2. Accuracy and Structuredness Results.

Qualitative analysis of student comments showed that overall students found the visualization useful in showing different mathematical themes from which a problem could be approached. One student indicated that the visualization was helpful when he could not solve a problem. Students also stated that, over time and with more contributors, the system would become increasingly valuable for studying purposes. Students also commented that they became more cognizant of the connections amongst mathematics ideas and themes. It is noteworthy that students gained awareness that one could discuss properties of math problems and their relevant themes rather than simply solve them.

For part A of the activity, video recordings revealed that students worked on the tagging individually but commented to each other about the connections they created in the group visualization (between the problem and their group category). In part B, the videos indicated that students made concerted efforts to solve problems collaboratively, although some groups were more successful than others. One group showed two students taking turns actively solving, while a third member interjected occasionally with valuable comments. Another group initially solved questions individually then compared answers as an approach; soon after it became clear that one member was faster at solving than the others and he took on more responsibility for this portion of the activity. Discussion in the groups ranged from reading out questions, talking aloud while thinking through strategies, asking other members what certain formulas are, to verifying or questioning approaches and answers. Since students only solved problems that were tagged as their group's category, students also discussed the appropriateness of the tag for certain questions. In the post-test, one student commented that solving math problems in groups was valuable because it is not something they usually get to do. In the last part of the activity, the amount of discussion varied amongst groups. Some discussion revolved around whether they solved the problem from the category's perspective to determine whether they agree with the connection or not. In some groups, the agreement was straightforward, and voting was completed without debate.

Study 2: Collaborative Problem-solving

The second study addressed our objective by engaging students in a slightly more complex collaborative problem-tagging activity where they worked individually and in groups to identify the important conceptual elements within a set of qualitative physics problems. Key differences between this and the math curriculum lies in students providing answers as well as tags in the first step, which gave groups more collective information to work with in subsequent steps. Also, with the problems being more conceptual in nature, students could focus on discussing higher-level issues rather than manipulating numbers, which seemed to elicit more individual work. Learning outcomes are measured in terms of i) students' precision in answering the qualitative problems, and ii) their classification and assignment of conceptual characteristics to the physics problems. The problem-tagging activity is followed by a problem set-up activity, where students worked collaboratively to set up the equations and approaches for solving a long-answer physics problem. The curriculum activity was designed for enactment over a class period of ninety minutes, as an end-of-term review activity. Students were randomly assigned to groups of four. Each student was given a laptop computer and the group was also provided with a projection screen to be used for the display of the collaborative visualizations.

Participants

A total of thirty-two student volunteers enrolled in grade twelve physics at our research partner high school participated in the activity. Two smart classroom sessions were conducted over two days with sixteen students in each cohort. The curriculum activity was co-designed with a high-school physics teacher.

Design

Part A - Individual Solving & Tagging

The curriculum was designed to engage several small groups of students working in parallel as they answered a common set of sixteen multiple-choice concept-based physics questions. Each student answered and tagged four multiple-choice concept problems with expert-defined "element" tags (e.g., Newton's first law, net force, kinetic energy, conservation of momentum, etc.), selected in advance by the co-design team. The goal of this activity was to familiarize students with the elements with which experts would typically categorize problems.

Part B - Group Review

Once students had completed tagging their four concept problems, they worked as a group to review the responses and tags made by other students for four of the problems. Collaborative visualizations displaying those results were generated to facilitate this process. Students were instructed to critique the various solutions contributed by their classmates, as well as the collective tags, then re-negotiate the "definitive" answers and element sets, and write a brief rationale to explain their choice of elements (Figure 3).

Part C - Long Problem Setup

Upon completion of the concept question reviews, four complex quantitative physics problems were presented to each group. For each long problem, students were asked to select from a list of four concept questions that they felt was most related to the long problem. Once the selection was made, students were asked to choose a set of elements and equations that would help set up the problem for solving. Finally, groups provided explanations for their choice of formulas.

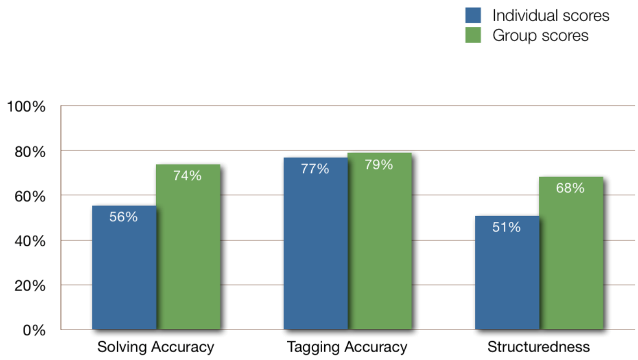


Figure 3. Students Reviewing Class Responses in Step 2. Figure 4. Accuracy and Structuredness Results.

Data Analysis

Data were gathered during the activity sessions. Preliminary data screening was conducted to determine whether there were differences between students who participated in the activity in day 1 compared to those who participated in day 2. Their individual scores (part A) as well as group scores (part B) were assessed and no significant differences were found. Participant data in from both days was pooled and analyzed together. Individual performance on concept questions was compared with group performance using a paired samples *t* test. Element tags were analyzed using accuracy and structuredness scores, using the protocol described in study 1. Qualitative analysis was performed on video captured and audio data collected at group workstations.

Results

A paired samples *t* test was conducted to evaluate whether students performed better at the group review step than at solving individually. The results indicate that the mean group scores (M = 74.00, SD = 23.78) were significantly higher than the mean individual scores (M = 55.65, SD = 30.40), $t(30) = -2.74$, $p < .05$ (see Figure 4). Further analyses showed no significant differences among group scores between the eight groups. For tagging accuracy and structuredness, groups tended to tag their problems closer to the expert model than individuals, with average accuracy scores of 79% (SD = 4.84) compared to 77% (SD = 7.40), although the difference was only marginally significant. In terms of structuredness, students' group scores (M = 68.43, SD = 13.21) were significantly higher than their individual scores (M = 50.92, SD = 18.51) by approximately 18%, $t(30) = -5.654$, $p < 0.05$ (see Figure 4).

Video recordings of classroom activity and group level audio data indicated that students generally worked independently in part A, with some seeking help with equations (e.g., for elastic potential energy) and a few asking group members how to approach certain problems. In part B, students noted what others in the class chose as the right answer and compared that to their own choices. They discussed formulas and verified which should be used, and used analogies to explain concepts to other group members. Students also discussed how the problem should be approached, often using their own answers as a starting point for discussions. They also seemed to connect element tags to formulas that should be used in solving the problems, which was in line with analyses of the written rationales.

Discussion

Collective Knowledge and Connections

Both studies employed large projected displays of the aggregated input from the individual members working within a small group. In study 1 (math curriculum), the collaborative visualizations served as representations of collective knowledge on small group levels as well as on a whole class level. They helped students make connections between math problems and themes, but also provided researchers a means of assessing the connections that students made by comparing their answers to those of the teacher (or another normative source). The collaborative visualizations also provided a record of the aggregated connections, artifacts (e.g. problems, solutions) and communications amongst students over time, which can be used to inform the design of subsequent learning activities. As a result, students in study 1 seemed to gain an appreciation that problems may exhibit characteristics from more than one distinct category. However, they did not discuss meaningful differences among concepts, perhaps due in part to the overall activity design and to the limited amount of information shared in the visualizations.

Study 2 (physics curriculum) also showed improved tagging accuracy over time when connecting problems to underlying concepts. In particular, the structuredness score revealed a significant difference between individual and group efforts, which indicates an improved willingness of students to characterize problems from different perspectives. This was accompanied by a significant improvement in accuracy for answering conceptual questions over the course of the activity session. From analysis of video and audio data, the improvement is likely due to a combination of the collaborative discourse and the "wisdom of the crowd" (i.e., by way of reviewing of class answers). For each question that groups reviewed, students took into consideration how fellow classmates answered the question and what element tags they assigned to it. However, the bulk of the group discourse was centered on the elements and how they contributed to the correct answer. Collaborative tagging of elements also seems to help students set up quantitative physics problems. Rationales for tag selections as well as their collaborative discourse around the elements appear to have provided a rich conceptual space for students in which they can organize key concepts.

Dynamic representations of the knowledge base were important design features in the smart classroom. In study 1, the four corners of the room represented different location of expertise, with the aggregated knowledge of all four groups shown at the front of the room. Students took notice of the representation and looked around the room at various points of the activity, paying special attention to the front of the room. In study 2, a more distributed approach was taken to represent the knowledge base. Class responses were associated with particular questions (e.g., in the review step, the group would see fellow classmates' responses for question one while reviewing the first question, once they submitted their final answers for the question, they would then review the next question and see the class responses pertaining only to the second question). In study 2, there were no persistent representations of the knowledge base, and the behaviors of the class as a whole was less cohesive. Students tended to look around at other groups for indications of their progress completion during the activity, rather than how other groups made connections and solved problems. Casual comparisons between student behaviors in study 1 and 2 indicate that specialization of groups and persistent representations of the class-wide knowledge base may improve learning experiences in the smart classroom (e.g., embodiment, engagement), although this was not explicitly measured in either study.

Orchestration of Smart Classroom Activities

Both studies also employed the notion of orchestration, which was in the form of an individual tagging phase, group review and a reflection phase. Asking students to perform slightly different tasks using collaboration scripts at each of these phases revealed how they can be utilized optimally for learning in the smart classroom. The individual phase in both studies served as a "model-building" phase of the participants' initial collective knowledge base. Although some students occasionally consulted with fellow group members during this stage, they generally completed the phase individually. The affordances of the smart classroom were not put to use in this phase in both instances, which means this phase can potentially be placed outside the room, perhaps as homework assignment.

Asking students to only tag concepts to problems in the first phase (as in study 1) provided limited information for groups to discuss in the second phase. Giving additional collective information to students in study 2 (i.e., more tags, adding answers to question) seemed to elicit more deliberation, but the goal of the discussions tended to be restricted to finding the correct answer. Perhaps if a richer context was used as part of the activity and a variety of information, including different types of connections and media, was given to students, the collaborative discourse could increase in complexity and raise the level of knowledge construction.

In the reflecting scripts, we expected to see more discourse around group decisions and creating rationales. In study 1, group members seemed to attain a similar mental model of each of the problems they were asked to reflect upon, either through directly solving the problem, or in discussing the strategy to solve the problem. Not much discussion was necessary to reach consensus. In study 2, discussions around problem-solving strategy were complementary to selecting element tags. One person tended to take over responsibility in creating the rationales based on what the group talked about. In order to draw out student thinking behind the reflection process, scripts may need to include more "think aloud" instructions for students to share their thoughts. Alternatively, we could recognize reflection as an individualistic process and script accordingly, rather than force this to be a collaborative process.

Recent research on scripting approaches has focused on "fading" such scaffolds over time (Wecker & Fischer, 2007), or providing scripts with adaptive scaffolds (Dillenbourg & Tchounikine, 2007; Rummel and Weinberger, 2008). While these investigations are still in their infancy, we aim to make use of our dynamically generated knowledge base to modify collaborative processes in real-time, which can provide opportunities for teachers and students to monitor, evaluate, and adapt learning activities during class time. For example, progress reports could be given to students and groups between activity phases, and the amount of scaffolding provided to groups may be adjusted based on formative assessments. Data may be provided to teachers in real-time, allowing them to guide students more effectively, and even potentially "flagging" those who need assistance most. Using more complex "intelligent agents" (e.g., teacher agents, student agents, and group agents) for data mining and reporting group activity, we wish to enhance the awareness and detection capabilities of the smart classroom framework to support group collaborations, student interactions, as well as classroom management. The present research represents an early effort in understanding how collaboration scripts organize smart classroom activities.

Collaborative Activities for Building Knowledge Communities

We recognize from prior research that a strong knowledge community can only be formed over a lengthy period of time (i.e., over weeks and months) around a shared set of knowledge and experiences, rather than in the short time frame in the enactment of the two studies described above. However, these short studies granted us the opportunity to look for evidence of how students might use real-time collective knowledge in various representations and how its use would form a cohesive community of learners. Examining student interactions from both studies, we gained valuable insight regarding the importance of i) representing clear, defined expertise or roles for students or groups of students to undertake; ii) providing clear and persistent representations of the community's knowledge base; and iii) focusing on a rich learning context to guide collaborations. Individuals and/or groups that assume distinct roles strengthen the collective knowledge base, possibly by imparting students with a sense of identity and ownership within the community. The presence of clear and persistent representations of the knowledge base provides students with concrete evidence of their shared knowledge, however dynamic the knowledge may be, and allow ideas to be built upon. Moving forward within the larger research program, our team will apply the findings from these initial studies and investigate more complex pedagogical configurations, involving longer duration curriculum and dynamic changes to the shared knowledge base over time. We will further develop the smart classroom environment and curriculum materials based on our ongoing series of design-oriented studies with a view to extend our basic approach to build knowledge communities for transformative learning.

References

Bereiter, C., & Scardamalia, M. (2003). Learning to work creatively with knowledge. In E. De Corte, L. Verschaffel, N. Entwistle, & J. van Merriënboer (Eds.), *Powerful learning environments: Unraveling basic components and dimensions* (Advances in learning and instruction series, pp. 55-68). Oxford, UK: Elsevier Science.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences*, *2*(2), 141-178.

Brown, A. L., & Campione, J. C. (1990). Communities of learning and thinking, or a context by any other name. In Kuhn, D., (Ed.), *Developmental Perspectives on Teaching and Learning Thinking Skills*, (vol 21, pp. 108-126). Basel, NY: Karger.

Brown, A.L., & Campione, J.C. (1996). Psychological theory and design of innovative learning environments: On procedures, principles, and systems. In L. Schauble & R. Glaser (Eds.), *Innovations in learning: New environments for education* (pp. 289-325). Mahwah, NJ: Lawrence Erlbaum Associates.

De Wever, B., Van Keer, H., Schellens, T., & Valcke, M. (2009). Structuring asynchronous discussion groups: the impact of role assignment and self-assessment on students' levels of knowledge construction through social negotiation. *Journal of Computer Assisted Learning*, 25, 177-188.

Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, *32*(1), 5-8.
Dillenbourg, P., & Tchounikine, P. (2007). Flexibility in macro-scripts for computer-supported collaborative learning. *Journal of Computer Assisted Learning*, *23*(1), 1-13.
Hasemann, K. & Mansfield, H. (1995). Concept mapping in research on mathematical knowledge development: Background, methods, findings, and conclusions. *Educational Studies in Mathematics*, *29*, 45-72.
King, A. (1999). Discourse patterns for mediating peer learning. In A. M. O'Donnell & A. King (Eds.), *Cognitive perspectives on peer learning* (pp. 87-117). Mahwah, NJ: Lawrence Erlbaum Associates.
Kobbe, L., Weinberger, A., Dillenbourg, P., Harrer, A., Hämäläinen, R., & Fischer, F. (2007). Specifying computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning*, *2*(2–3), 211–224.
Kollar, I., Fischer, F., & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning & Instruction*, *17*(6), 708-721.
Linn, M. C., & Eylon, & B. - S. (2006). Science education: Integrating views of learning and instruction. In P. A. Alexander, & P. H. Winne (Eds.), *Handbook of Educational Psychology* (pp. 511–544). Mahwah, NJ: Lawrence Erlbaum Associates.
Partnership for 21st Century Skills. (2009, December). *Framework for 21st Century Learning*. Retrieved October 15, 2010, from http://www.21stcenturyskills.org/documents/P21_Framework.pdf
Penuel, W. R., Roschelle, J. & Shechtman, N. (2007). Designing formative assessment software with teachers: An analysis of the co-design process. *Research and Practice in Technology Enhanced Learning*, *2*(1), 51-74.
Peters, V. L., & Slotta, J. D. (2009). Co-designing curricula to promote collaborative knowledge construction in secondary school science. In C. O'Malley, D. Suthers, P. Reimann, & A. Dimitracopoulou (Eds.), *Proceedings of the Computer Supported Collaborative Learning (CSCL) Conference* (pp. 204-213). International Society of the Learning Sciences (ISLS).
Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., Kyza, E., Edelson, D., & Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. *Journal of the Learning Sciences*, *13*(3), 337-386.
Rummel, N., & Spada, H. (2005). Learning to collaborate: An instructional approach to promoting collaborative problem-solving in computer-mediated settings. *Journal of the Learning Sciences*, *14*(2), 201-241.
Rummel, N., & Spada, H. (2007). Can people learn computer-mediated collaboration by following a script? In F. Fischer, H. Mandl, J. Haake, & I. Kollar (Eds.), *Scripting computer-supported communication of knowledge: Cognitive, computational and educational perspectives* (pp. 39–55). New York: Springer.
Rummel, N., & Weinberger, A. (2008). New challenges in CSCL: Towards adaptive script support. In G. Kanselaar, V. Jonker, P.A. Kirschner, & F. Prins, (Eds.), *International perspectives of the learning sciences: Cre8ing a learning world. Proceedings of the 8th international conference of the learning sciences* (pp. 338–345). Utrecht, NL.
Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *The Journal of the Learning Sciences*, *3*(3), 265-283.
Scardamalia, M., & Bereiter, C. (1996). Computer support for knowledge-building communities. In T. Kotchmann (Ed.), *CSCL: Theory and practice of an emerging paradigm.* Mahwah, NJ: Lawrence Erlbaum Associates.
Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (pp. 97-118). New York: Cambridge University Press.
Slotta, J. D. (2010). Evolving the classrooms of the future: The interplay of pedagogy, technology and community. In Mäkitalo-Siegl, K., Kaplan, F., Zottmann, J. & Fischer, F. (Eds.). *Classroom of the Future: Orchestrating collaborative spaces* (pp. 215-242). Rotterdam: Sense.
Slotta, J. D., & Peters, V. L. (2008). A blended model for knowledge communities: Embedding scaffolded inquiry. *Proceedings of the 8th international conference on international conference for the learning sciences*, 2, 343-350.
Stegmann, K., Weinberger, A., and Fischer, F. (2007). Facilitating argumentative knowledge construction with computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning, 2*, 421-447.
Wecker, C., & Fischer, F. (2007). Fading scripts in computer-supported collaborative learning: The role of distributed monitoring. In C. A. Chinn, G. Erkens, & S. Puntambekar (Eds.), *Proceedings of CSCL 2007: Of mice, minds and society* (pp. 763–771). New Brunswick, NJ, USA.
Weinberger, A., Ertl, B., Fischer, F., & Mandl, H. (2005). Epistemic and social scripts in computer-supported collaborative learning. *Instructional Science, 33*, 1-30.

ACODEA: A Framework for the Development of Classification Schemes for Automatic Classifications of Online Discussions

Jin Mu, LMU München, Leopoldstrasse 13. 80802 München, GERMANY, jin.mu@psy.lmu.de
Karsten Stegmann, LMU München, Leopoldstrasse 13. 80802 München, GERMANY, karsten.stegmann@psy.lmu.de
Elijah Mayfield, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA, elijah@cmu.edu
Carolyn Rosé, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA, cprose@cs.cmu.edu
Frank Fischer, LMU München, Leopoldstrasse 13. 80802 München, GERMANY, frank.fischer@psy.lmu.de

Abstract: Research related to online discussions frequently faces the problem of analyzing huge corpora. Natural Language Processing technologies may allow automating the analysis. However, the state-of-the-art in machine learning and text mining approaches yields models that do not transfer well between corpora related to different topics. Also segmenting is a necessary step, but frequently, trained models are very sensitive to the particulars of the segmentation that was used when the model was trained. Therefore, in prior published research on text classification in a CSCL context, the data has been segmented by hand. We discuss work towards overcoming these challenges. We present a framework for developing coding schemes optimized for automatic segmentation and topic independent coding that builds on this segmentation. Our results show that our coding scheme can be fully automated by using a tool called SIDE. Finally, we discuss how fully automated analysis can enable context-sensitive support for collaborative learning.

Why should Online Discussions be Coded Automatically?

Online discussions have been widely used in the field of CSCL to foster collaborative knowledge construction. Learners work together to exchange ideas, negotiate meaning and formulate understanding (De Laat and Lally 2003). One important feature of online discussions is that this kind of communication produces a huge body of digital data as a byproduct of the interaction. Researchers are therefore confronted with the opportunity as well as the challenge of analyzing online discussions at multiple levels, such as quality of argumentation, or social modes of interaction (e.g., Weinberger & Fischer, 2006). A variety of multidimensional frameworks have been employed (cf. Clark et al., 2007). In this study, we focus specifically on analysis of what has previously been called micro-argumentation (Weinberger & Fischer, 2006), with the idea of expanding to other dimensions of analysis in future work.

Evaluation of discussion quality consumes a huge amount of resources in research projects related to online discussions. In order to address this problem, Rosé and colleagues (2008) reported a strand of experimental studies with about 250 online discussions (cf. Stegmann, Weinberger & Fischer, 2007; Weinberger et al., 2005; Weinberger, Stegmann & Fischer, 2010) where about 25% of all human resources in the research project were spent to analyze online discussions on multiple dimensions. Human coders had to be trained to annotate segments of these data using a multi-dimensional coding scheme that operationalized aspects of content as well as manner of argumentation and social modes of interaction. While uncovering findings related to how group knowledge construction works often make those efforts worth the time and energy they require, analyzing a huge body of online discussions by hand is an arduous task that slows down the progress of the research substantially. An automatic and thus faster classification of online discussions may affect the whole research process positively. One possible impact may be that an increasing number of researchers may be willing to analyze online discussions on multiple dimensions. Moreover, a part of the freed-up resources may be used to conduct follow-up studies or to try out pioneering approaches to data analysis.

Automatic classification may not only facilitate research on online discussions. Automatic classification also allows for adaptive collaborative learning support (ACLS; Kumar et al., 2007; Kumar & Rosé, in press; Walker, Rummel & Koedinger, 2009) to foster the quality of collaborative knowledge construction during online discussions (e.g. Gweon et al., 2006; Walker, Rummel & Koedinger, 2009). Online discussions could be analyzed in real-time and instructional support measures like hints or scaffolds could be adapted to the quality of certain aspects of the collaboration. For example, learners who are unable to provide warrants and grounds for their claims may get offered a scaffold to construct better arguments. Learners who fail to relate their contribution to other learning partners may be explicitly asked to provide feedback. Against this background, we report a use case on the application of a tool for Natural Language Processing (SIDE) in a multi-layer framework which has optimized for fully automatic segmenting and context independent coding. We begin by providing an overview of the state-of-the-art in the application of NLP technologies in CSCL research.

Applying NLP Technologies in CSCL

Natural Language Processing has long been used to automatically analyze textual data. The need for involving technology from NLP in the process of content analysis is growing in the presence of the Web and distance learning (Duwairi, 2006). For instance, the NLP methods of content analysis have been developed for the automatic grading of essays (Duwairi, 2006; Landauer, Laham, & Foltz, 2003;); and for the intelligent and cognitive tutoring (Rosé & VanLehn, 2005; Diziol, Walker, Rummel & Koedinger, 2010).

In the last few years, researchers have begun to investigate various text classification methods to help instructors and administrators to improve computer supported learning environments (Kumar & Rosé, in press; Romero & Ventura, 2006). Text classification, is an application of machine learning technology to a structured representation of text, which has been a major focus of research in the field of NLP during the past decade. Typically, text classification is the automatic extraction of interesting, frequently implicit, patterns within large data collections (Klosgen & Zytkow, 2002). Nowadays, text classification tools are normally designed mainly for power and flexibility instead of simplicity (Romero & Ventura, 2006), which can assess student's learning performance, examine learning behaviour, and provide feedback based on the assessment (Castro, Vellido, Nebot, & Minguillon, 2005). Consequently, most of the current text classification tools are too complex for educators to use, and thus their features go well beyond the scope of what an educator might require (Romero & Ventura, 2006).

Therefore, TagHelper (Rosé et al., 2008) and its successor SIDE (Mayfield & Rosé, 2010) were developed to automate the content analysis of collaborative online discussions. As a publically available tool, TagHelper has been downloaded thousands of times in over 70 countries. Recently, application of TagHelper for automated tutoring and adaptive collaboration scripts have been extensively researched (Kumar & Rosé, in press). In order to make TagHelper tools accessible to the widest possible user base, default behaviour has been set up in such a way that users are only required to provide examples of annotated data along with un-annotated data. TagHelper first extracts features like line length, unigrams, bigrams and part-of-speech bigrams from the annotated data. An interface for constructing rule-based features is also provided. In SIDE, more sophisticated support for feature construction is included, such as regular expressions, which are important in the area of information extraction and named entity recognition, which we make use of in the study reported in this paper. Recent work has also yielded approaches for automatic feature construction (Mayfield & Rosé, 2010b), which further enhances the ability to construct richer and more effective representations of text in preparation for machine learning. Tools such as TagHelper and SIDE then build models based on the annotated examples that it can then apply to the un-annotated examples. To get the best results, both tools allow users to switch easily between different machine learning algorithms provided by Weka (Witten & Frank, 2005). Namely, they are for example Naïve Bayes, SMO, and J48.

Despite the effectiveness of applying TagHelper to analyze text-based online discussions, at least two challenges associated with the current NLP approach are still needed to be addressed. First, the automatic approach has so far only been demonstrated on annotated examples from corpora that come from a single scenario, and the generated model is quite context sensitive and case dependent, and has not been demonstrated to transfer well to online discussions with different topics. Second, if no natural borders of segments are provided to the tool, the automatic segmentation that it has been able to get prior to the work reported here has not been satisfactory (Rosé et al., 2008). However, such an automatic segmenting is imperative for us to investigate the real-time adaptive fading, which is only possible when segmentation is also done automatically. Besides the shortcomings of applying TagHelper into the field of CSCL mentioned above, the accuracy and reliability is often not as optimal as we would like it to be, especially when the segmentation is done automatically (Rose et al., 2008). This motivates our investigation to explore whether the use of more advanced natural language processing technology can offer fully automatic and context independent automating techniques for content analysis.

In this paper we explore several enhancements to this technology in order to overcome these challenges. For example, one promising direction to consider is the integration of information extraction techniques for improving content analysis. Previous work on applying NLP in the field of CSCL, generally accepted raw text as input for segmenting and coding, and the features used for classification were very low-level and simplistic. The new approaches used in the current study have strengths in information extraction, which allow the construction of a more sophisticated representation of the text to build the classification models on. Such technology includes named entity recognition, which is an active area of research in the field of language technologies.

Part-of-Speech tagging (PoS) is the process of assigning a syntactic class marker to each word in a text (Brill, 1992; Mora & Sánchez Peiró, 2007), therefore, a PoS tagger can be considered as a translator between two languages: the original language that has to be tagged and a "machine friendly" language formed by the corresponding syntactic tags, such as noun or verb. As Poel et. al.(2007) proposed, PoS tagging is often only one step in a text processing application. The tagged text could be used for deeper analysis. Instead of using PoS as the default generalized features, it makes sense to apply modified and specialized PoS categories and thereby to facilitate automatic segmentation if the unit of analysis is syntactically meaningful.

The goal of *Named Entity Recognition* (NER) is to classify all members of certain categories of "proper names" appearing in the raw text, into one of seven categories: person, organization, location, date, time, percentage, and monetary amount (MUC6, 1995). Core aspects of NER are entity and mentions. Mentions are specific instances of entities. For example, mentions of the entity class "location" are New Brunswick, Rhodes, and Hong Kong. Therefore NER provides not only additional features based on extracted entities for each word, but also a more context-independent way to train automatic classifiers. The mentions of New Brunswick, Rhodes, Hong Kong are cities in, for example, the discussions about the past and upcoming three CSCL conferences, while Bloomington, Utrecht, and Chicago would have the same semantic function within discussions about the past three ICLS conferences. As an initial step of pre-processing in information extraction applications, an automatic classifier that had been trained with predefined entities (e.g. "location") instead of specific mentions (e.g. Hong Kong) might have more flexibility for modelling contextual information, potentially improving classification performance. More recently, there have been tasks developed to deal with different practical problems (IREX and CoNLL-2002), in which every word in a document must be classified into one of an extensive set of predefined categories, rather than only identifying names of people and locations.

With the support of current approaches in information extraction, the input to SIDE is assumed to be enhanced in a fully automatic way to be less context dependent. In the following section, we will present the multi-layer framework for the development of classification schemes for automatic segmentation and coding.

The Automatic Classification of Online Discussions with Extracted Attributes (ACODEA) Framework

Typical text classification for online discussions in CSCL is made to be applied by humans. They rely strongly on implicit knowledge held by human coders (e.g., understanding sentences with misspelled words or wrong grammar) to reach an acceptable level of reliability. Text classification that should be applied automatically has to account for the more limited features that are usually used to train automatic classifiers. Our following framework may support the development of such classification schemes.

Before delving into the specific processes of how the machine learning tool operates, we further clarify the concepts that are to be learned by SIDE. Witten and Frank (2005) details how data can be associated with classes or concepts, which should be reproducible by SIDE, intelligible to human analysts, and operational to be applied to actual examples. The starting point for understanding online discussion analysis is to define the coding schemas. In choosing the coding schemas, the researcher needs to determine what sized segments (which range from single word, sentence, paragraph, to the entire message) match with the desired and target activities to be coded (Strijbos et al., 2006). Thus the first target concept to learn is to classify, at each word, whether a segment boundary occurs. Similar to an earlier segmentation approach (Rosé et al., 2008), the concept of segmentation is implemented as a "sliding window" consisting of a specific number of words. In this way, any segmentation is possible since the boundary between any neighboring pair of words is a possible site for a segment boundary. The second concept considered here is to sort each unit of analysis (segment) to one or more categories (dimensions of analysis). For instance, a specific sentence, utterance or message is classified according to quality of argumentation or social mode of interaction (cf. Weinberger & Fischer, 2006).

Each individual instance (word in the text to be segmented and then coded) provides an input to machine learning, which is characterized by a fixed and predefined set of features or attributes. Text classification often requires data transforming into appropriate forms (Han & Kamber, 2006). Attribute construction (or feature construction), where new attributes are constructed and added from the given set of attributes, can help provide richer, more effective features for representing the text prior to text classification, consequently, ease the training of automatic classifiers as well.

Here we present our architecture for extracting features from the text in order to construct a representation suitable for applying machine learning to, either for the segmenting layer or the coding layer. The basic rules are to apply part-of-speech tagging and named entity recognition to extract features that are abstract enough to make interesting patterns apparent to machine learning algorithms and yield models that generalize well. On both the syntactic and semantic levels, rather than use predefined categories, we design customized sets of labels which extract information about the specific tasks or target activities we wish to classify. These labels align with behaviors which participants are expected to do during the discourse. In addition, the entire architecture is structured to cascade from one layer to the next, incorporating information from the previous layers to improve the current classifier's performance. Extracting attributes on the syntactic level benefits from the use of off-the-shelf grammatical part of speech taggers, while the layer related to semantic representation benefits from the inclusion of named entities and techniques from information extraction. The outputs of these layers is used as the attributes for the final classification layers of segmenting and coding. In this paper we want to explore the positive impact on performance of adding in abstract syntactic and semantic features above baseline feature spaces consisting of word-level representations such as unigrams and bigrams.

The automatic framework to analyze the content of online discussion can be further illustrate in the detailed flow processes below. At first, each single word in the raw data for training must be pre-processed by human coders to extract the syntactic and semantic features. These annotated examples, which reach acceptable reliability, can then be used to train classifiers for all defined ontologies. On the first layer, the part-of-speech tagger and named entity recognition are both applied independently. Secondly, human coders have to classify the borders between the segments. These human coded examples are used to train the automatic segmentation. However, the input to the segmentation classifier would be the preprocessed concepts from the syntactic attributes, instead of the raw text. After successful training of automatic classifiers for segmentation (i.e. high reliability regarding the identification of borders between segments), an automatic classifier segments all the data. This is required to provide equal training material for humans and automatic classifiers. Once the data is segmented, human coders have to classify all segments in the training data regarding the dimensions defined for the coding layer. These classifications are used to train classifiers for all defined dimensions. Again, the concepts from the semantic attributes are used to train the automatic classifiers. Finally, evidence is required that the automatic classifier (classifying and training with concepts from the predefined attributes) is reliable compared with human coders (classifying and training with raw data). This is needed because the whole procedure has one major risk: Several classifications are made in a row, and thus errors on the different layers are cascaded. Therefore, the final automatic classification must ultimately be checked against pure human coding.

Research Questions

In the following, we will present a use case of the multi-layer framework. Our research questions are:

RQ1: Can the automatic classification of attributes be trained reliably compared with human coding?
RQ2: Can the automatic segmentation be implemented reliably compared with human segmenting?
RQ3: Can the framework be used to train context independent segmentation with sufficient reliability?
RQ4: Can the automatic coding be implemented reliably compared with human coding?
RQ5: Can the framework be used to train context independent classification with sufficient reliability?

Methods

Participants and Learning Task

Eighty-four (84) students of Educational Science at the University of Munich participated in this study. Students were randomly assigned to groups of three. Each group was randomly assigned to one of three experimental conditions. The task of the groups was to join a collaborative, argumentative online discussion and solve five case-based problems with the help of an educational theory. The computer-based learning environment used in this experiment is a modified version of the one employed by Stegmann, Weinberger and Fischer (2007).

The subject of the learning environment is Weiner's attribution theory (1985) and its application in education. The students read the text of attribution theory and the text of introducing argumentation individually. In the collaborative learning phase, three problem cases from practical contexts were used as a basis for online discussions. The case "Math" describes the attributions of a student with respect to his poor performance in mathematics. In the case "Class reunion" a math tutor talks about how he tries to help female students deal with success and failure in assignments. The case "Asia" describes differences in school performance between Asian and American/European students that were explained by the attribution theory. Another two cases were used in the pre and post test, which mainly concern the factors which affect a student's "Major choice" at the university and student's explanation for failure in the exam of "Text analysis".

Data Source and Procedure

We collected 140 conversation transcripts, each of which contained the full interaction from one group, and was targeted to a single scenario. Altogether, there are 74764 words in the corpus. Two human coders analyzed almost one tenth of the raw data (distributed over five cases), which have been further used for training the customized algorithms for the classification on the extracing attributes, segmenting and coding layers by SIDE (Mayfield & Rosé, 2010). SIDE is the successor of TagHelper (Rosé et al., 2008) and includes an annotation interface allowing for automatic and semi-automatic coding more easily. To train such classifiers with SIDE we had to provide examples of annotated data. SIDE extracted multiple features from the raw data, like line length, unigrams, bigrams, Part-of-speech bigrams, etc. Machine learning algorithms used these features to learn how to classify new data. As output, SIDE builds a model based on the human annotated data. This model can then be easily applied to classify un-annotated data, and then the assigned codes can be further reviewed on the annotation interface by modifying the codes the human coders disagree with. Furthermore, SIDE employs a consistent evaluation methodology referred to as 10-fold cross-validation, where the data for training the models can be randomly distributed into 10 piles. Nine piles are combined to train a model. One pile is used to test the model. Self-reported reliability is calculated by averaging across the ten iterations.

Statistical Tests

The reliability of the coding was measured using Cohen's Kappa value and percent accuracy. Both of them have been regarded as accepted standards for measuring coding reliability. The criterion for success is reaching a level of agreement with a gold standard as measured by Cohen's Kappa that is .7 or higher (Strijbos et al., 2006).Here it is worthwhile to further clarify that the present study was undertaken to evaluate different types of Kappa in the distinguishable phases, including (1) inter-rater agreement between human coders to evidence the reliability of training examples; (2) internal generated Kappa by the 10-fold cross-validation to certify the reliability of the SIDE training models, and (3) the conclusive Kappa between human coders and SIDE.

Application of the Framework

(i) Our coding layer was defined with respect to the approach of argumentative knowledge construction. Learners construct arguments in interaction with their learning partners in order to acquire knowledge about argumentation as well as knowledge of the content under consideration (Andriessen, Baker, & Suthers, 2003). Therefore on this layer we are mainly concerned with the following categories, based on the micro-argumentation dimension of the multidimenional framework generated by Weinberger and Fischer (2006): (a) *Claim* is a statement that advances the position learners take to analyze case with attribution theory. (b) *Ground* is the evidence from case to support claim. (c) *Warrant* is the logical connection between the grounds and claims that present the theoretical reason why a claim is valid. Consequently, (d) the *Inadequate claim* should be differentiated in the coding, which concerns other related educational theory to explain case. (e) *Evaluation* is the agreement with learning partner or not. There are more dimensions to indicate the (f) *Empty Message* and (g) *Scripts*, both of them are the computer generated to structure the argumentative discourse, and finally (i) *Others*, which cannot be sorted by any other dimensions.

(ii) The unit of analysis was defined as a sentence or part of a compound sentence that can be regarded as 'syntactically meaningful in structure, regardless of the meaning of the coding categories' (cf. Strijbos et al., 2006, p. 37). For instance, according to these rules of segmentation, punctuation and the special words like 'and' are boundaries to segment compound sentences if the parts before and after the bounder are 'syntactically meaningful' segments. This size of segment has been proved to be reliable (Strijbos et al., 2006), and suitable for the coding dimension conducted in the current study.

(iiia) Regarding the syntactic attributes: The set of syntactic annotation consists of 12 different tags, which denote the grammatical features of a word class. Each word in the computerized data can be pre-processed into multiple and syntactic categories. An example of such a tag is: Term, Verb, Property, Conjunction, Comma/Stop Symbol, and so on. These tags are a reduced version of the full tag set, making it more suitable for machine learning. Some stop words like Pronoun are clustered into the class of Others. For example, "The math-failure of the son is external stable, because the entire family is not good at math." can be replaced into "Others / Terms / Others / Others / Terms / Verbs / Property / Property /CommaSymbol /Others / Others / Property / Terms / Verbs / Others / Property / Others / Terms /StopSymbol /".

(iiib) Regarding the semantic attributes, each single word in the text can fall into one of the multiple categories, either (a) *Case*, key words from problem space, (b) *Theory*, key words from the concerned conceptual space (actually, attribution theory in the present study), or (c) *Extraneous theory*, from the related educational theory. In addition, there are words that are important in reflecting the (d) *evaluation* either positive or Negative among partners (which refers to key indicator of Counterargument), (e) *Empty Message*, and even (f) *Others activities*, can be extracted in this phase. Therefore the mentioned above example can be represented on this layer as Others (The) / Case (math-failure) / Others (of) / Others (the) /Case (son) / Others (is) /Theory (external) /Theory (stable) /Others (,) /Others (because) /Others (the) / Others (entire) /Case (family)/Others (is) /Others (not) /Others (good) /Others (at) /Case (math) / Others (,). All of the categories are choosen because they might support the coding on the classification layer. For instance, according to our learning task a claim would typically contain both case and theory information, while a ground mainly includes case information, and a warrant only includes elaborations on the attribution theory.

Results

Two coders created the training material for SIDE. The inter-rater agreement between two human coders was Cohen´s Kappa =.93 on the syntactic-attributes layer and Cohen´s Kappa = .97 on the semantic-attributes layer. We achieved a high value of Cohen's Kappa = .96 for the segmentation layer and Cohen's Kappa = .71 for the coding layer. These results indicate acceptable human baseline performances for SIDE to be trained to analyze the un-annotated data regarding the extracting attributes, segmentting and coding layers.

RQ1: Training of the Layer of Extracting Attributes

SIDE achieved an average Cohen's Kappa = .94 (accuracy = 91.7%) with the training material on the syntactic layer, and an average of Cohen's Kappa = .93 (accuracy = 96.0%) on the semantic layer. A human coder and SIDE achieved an agreement of Cohen's Kappa = .89 (accuracy = 92.0%) on the syntactic layer, and Cohen's

Kappa = .94 (accuracy = 97.4%) on the semantic layer. As shown in Table 1, the reliability of SIDE to analyze text on the syntactic and semantic layers is satisfactory across all 5 cases. Because the precision on the layer of extracting attributes greatly influences the performance of the steps further in the chain of linguistic treatments, the accuracy of the PoS tagger and named entity recognition is very important.

Table 1: Reliability of Automatic attributes extraction within 5 Cases (SIDE vs. Human).

Case	Syntactic-attributes Layer		Semantic-attributes Layer	
	Cohen's Kappa	Accuracy	Cohen's Kappa	Accuracy
Major choice	0.90	92.2%	0.95	98.1%
Math	0.85	88.1%	0.92	96.9%
Class reunion	0.91	92.9%	0.90	96.0%
Asia	0.93	94.4%	0.96	98.3%
Text analysis	0.88	90.3%	0.94	97.7%

RQ2 & RQ3: Training of the Segmentation Layer

Table 2: Comparison without and with the layer of extracting attributes to automate the content analysis (SIDE vs. Human).

	Without extracting attributes		With extracting attributes	
Segmentation	Cohen's Kappa	Accuracy	Cohen's Kappa	Accuracy
Internal SIDE Kappa	0.84	96.7%	0.98	99.6%
Kappa between SIDE and Human	0.86	97.0%	0.97	99.3%
Major choice	0.80	96.7%	0.95	99.1%
Math	0.86	96.6%	0.96	98.9%
Class reunion	0.87	97.0%	0.97	99.3%
Asia	0.90	97.7%	0.99	99.7%
Text-analysis	0.83	96.9%	0.98	99.6%
Classification	Cohen's Kappa	Accuracy	Cohen's Kappa	Accuracy
Internal SIDE Kappa	0.70	75.6%	0.77	81.3%
Kappa between SIDE and Human	0.61	67.8%	0.81	84.5%
Major choice	0.63	71.2%	0.77	82.9%
Math	0.67	72.3%	0.78	82.6%
Class reunion	0.47	58.5%	0.76	81.0%
Asia	0.53	63.1%	0.85	87.5%
Text-analysis	0.68	75.0%	0.87	89.2%

Internal Cohen's Kappa = .98 (accuracy = 99.6%) was achieved by SIDE when it attempted to automatically segment the text, when the raw text has been pre-processed to extract syntactic attributes. A human coder and SIDE achieved an agreement of Cohen's Kappa = .97 (accuracy = 99.3%). In addition, the inter-rater reliability within the five different cases is displayed in the Table 2. The algorithm for segmenting generated on the base of the layer of extracting attributes achieved sufficiently high Cohen's Kappa across the tested cases.

RQ4 & RQ5: Training of the Classification Layer

SIDE achieved an internal Cohen's Kappa = .77 (accuracy = 81.3%) using the extracted semantic attributes across all cases during training. The reliability across all cases comparing SIDE with a human coder (based on raw text) was sufficiently high (Cohen's Kappa = .81; accuracy = 84.5%). Table 2 shows the results within five different cases. Sufficient Cohen's Kappa values were achieved for all of the cases.

In summary, a domain-independent framework consisting of multiple layers is presented, which has been used successfully for the design, implementation and evaluation of a methodology for automatic classification of a large German text corpora. The reliability of automatic segmentation and coding has been reported across 5 cases with sufficient Cohen´s Kappa and accuracy. Compared with previous work, in which the raw data was directly used for the segmentation and classification, SIDE have been demonstrated to yield improved performance with the pre-processed layers to extract external attributes.

Discussion

The performance of SIDE to automate the content analysis has been demonstrated to be enhanced, by applying the multi-framework of the Automatic Classification of Online Discussions with Extracted Attributes (ACODEA). However, the language processing framework embedded in the argumentative knowledge construction and specific thematic domains is not usually designed to analyze other learning activities, such as problem solving or thought-provoking questioning. Depending on the domain as well as the type of target learning process, different sets of categories for the layers of extracting attributes and coding may be used for maximum performance. It remains to be seen whether such an automatic approach aimed at case independence will be further transferred to be able to analyze other collaborative activities, such as social interactivity and so on.

So far, the encouraging results indicate that it is possible to reach our ultimate goal of realizing adaptive collaboration scripts. In the proposed framework, the raw text was annotated using the techniques of part-of-speech tagging and named entity recognition before segmenting and classifying on the desired dimentions. The benefit is twofold. On one hand, the most challenging methodological problems have been successfully resolved, and the analysis of online discussion in the concerned domain has been evidenced to be fully automatic and context independent. On the other hand, one disadvantage of this information extraction is a vast reduction in the stored information, potentially losing some valuable features from the raw data that we had ignored on the layer of extracting attributes.

In addition, one interesting issue should be nevertheless still investigated to advance SIDE. It would be useful to be able to assess how "correct" the argumentation is, rather than only how complete the argumentation is, as we have done so far. From an epistemic perspective, an appropriate argument is more than a simple pile-up of information from problem and conceptual space, which includes a structurally appropriate connective between specific case and concerned theory. One possibility is that in the pre-processing step, the keywords from case and theory, which are correctly connected corresponding to an expert model, can be weighed automatically. This way, scaffolds provided by an adaptive collaboration script assisted by the automated and customized approach of qualitative content analysis can be much more powerful in its facilitation role, supporting valuable learning processes.

References

Andriessen, J., Baker, M., & Suthers, D. (2003). Argumentation, computer support, and the educational context of confronting cognitions. In J. Andriessen, M. Baker & D. Suthers (Eds.), *Arguing to learn: confronting cognitions in computer-supported collaborative learning environments* (Vol. 1, pp. 1-25). Dordrecht: Kluwer Academic Publishers.

Brill, E. (1992). A Simple Rule-Based Part-of-speech Tagger. *Proceedings of the Third Conference on Applied Natural Language Processing*. ANLP.

Castro, F., Vellido, A., Nebot, A., Minguillon, J., 2005. Detecting atypical student behaviour on an e-learning system. *In Simposio Nacional de Tecnologas de la Informacin y las Comunicaciones en la Educacion*, Granada, Spain (pp. 153–160).

Clark, D. B., Sampson, V., Weinberger, A., & Erkens, G. (2007). Analytic frameworks for assessing dialogic argumentation in online learning environments. *Educational Psychology Review, 19*(3), 343-374.

De Laat, M. F., & Lally, V. (2003). Complexity, theory and praxis: Researching collaborative learning and tutoring processes in a networked learning community. *Instructional Science*, *31*, 7–39.

Diziol, D.; Walker, E.; Rummel, N.; & Koedinger, K. R. (2010). Using intelligent tutor technology to implement adaptive support for student collaboration. *Educational Psychology Review*, *22*(1), 89-102.

Duwairi, R. M. (2006). A framework for the computerized assessment of university student essays. Computers in Human Behavior, *22*, 381–388.

Foltz, P. W., Laham, D., & Landauer, T. K. (1999). Automated essay scoring: Applications to educational technology. *In Proceedings of EdMedia*. .

Gweon, G., Rose, C. P., Zaiss, Z., & Carey, R. (2006). Providing Support for Adaptive Scripting in an On-Line Collaborative Learning Environment, *Proceedings of CHI 06: ACM conference on human factors in computer systems*. New York: ACM Press.

Han, J. & Kamber, M. (2006) *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers.

Klosgen, W., & Zytkow, J. (2002). *Handbook of data mining and knowledge discovery.* New York: Oxford University Press.

Kumar, R. & Rosé, C. P. (in press). Architecture for building Conversational Agents that support Collaborative Learning, *IEEE Transactions on Learning Technologies special issue on Intelligent and Innovative Support Systems for Computer Supported Collaborative Learning*.

Kumar, R., Rosé, C. P., Wang, Y. C., Joshi, M., Robinson, A. (2007). Tutorial Dialogue as Adaptive Collaborative Learning Support, *Proceedings of Artificial Intelligence in Education*.

Landauer, T. K., Laham, D., & Foltz, P. (2003). Automatic essay assessment. *Assessment in Education: Principles Plociy and Practice*, *10*(3).

Mayfield, E. & Rosé, C. P. (2010). An Interactive Tool for Supporting Error Analysis for Text Mining. *Proceedings of the North American Chapter of the Association for Computational Linguistics*.

Mayfield, E. & Rosé, C. P. (2010b). Using Feature Construction to Avoid Large Feature Spaces in Text Classification, *Proceedings of the Genetic and Evolutionary Computation Conference*.

Mora, G. & Sánchez Peiró, J. A. (2007). Part-of-Speech Tagging Based on Machine Translation Techniques. *Proceedings of the 3rd Iberian conference on Pattern Recognition and Image Analysis, Part I* (pp. 257 -264). Girona, Spain.

MUC6. Morgan Kaufmann Publishers, Inc. Proceedings of the Sixth Message Understanding Conference (MUC-6). Columbia, Maryland. 1995.

Poel, M., Stegeman, L. & op den Akker, H. J. A. (2007). A support vector machine approach to dutch partof-speech tagging. In M.R. Berthold, J. Shawe-Taylor, and N. Lavrac(Eds), *Proceedings of the 7th International Symposium on Intelligent Data Analysis* (pp. 274–283.) Springer Verlag, Berlin.

Romero, C., & Ventura, S. (2006). *Data mining in e-learning*, Southampton, UK: Wit Press.

Rosé, C. P., & VanLehn, K. (2005). An Evaluation of a Hybrid Language Understanding Approach for Robust Selection of Tutoring Goals, *International Journal of AI in Education*,*15*(4).

Rosé, C. P., Wang, Y.C., Cui, Y., Arguello, J., Fischer, F., Weinberger, A., & Stegmann, K. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer Supported Collaborative Learning 3*(3), 237-271

Stegmann, K., Weinberger, A., & Fischer, F. (2007). Facilitating argumentative knowledge construction with computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning*, *2*(4), 421-447.

Strijbos, J-W, Martens, R. L., Prins, F. J., Jochems, W.M.G. (2006). Content analysis: What are they talking about? *Computers & Education*, *46*, 29–48

Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge Universty Press.

Walker, E., Rummel, N., & Koedinger, K. R. (2009). CTRL: A research framework for providing adaptive collaborative learning support. *User Modeling and User-Adapted Interaction*, *19*(5), 387-431.

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computersupported collaborative learning. *Computers & Education*, *46*(1), 71-95.

Weinberger, A., Reiserer, M., Ertl, B., Fischer, F., & Mandl, H. (2005). Facilitating computer-supported collaborative learning with cooperation scripts. In R. Bromme, F. W. Hesse & H. Spada (Eds.), *Barriers and Biases in netwerk-based knowledge communication in groups*. Dordrecht: Kluwer.

Weinberger, A., Stegmann, K., & Fischer, F. (2010). Learning to argue online: Scripted groups surpass individuals (unscripted groups do not). *Computers in Human Behavior*, *26*, 506–515.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, *92*, 548-573.

Witten, I. H. & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier: San Francisco

Rating Dimensions of Collaboration Quality in Synchronous Collaborating Dyads: Findings and Interpretations

Georgios Kahrimanis, Irene-Angelica Chounta, Nikolaos Avouris, Human-Computer Interaction Group, Dept. of Electrical and Computer Engineering, University of Patras, Greece
Email: kahrimanis@ece.upatras.gr, houren@ece.upatras.gr, avouris@ece.upatras.gr

Abstract: Analysis and evaluation of CSCL activities are valuable for both building new knowledge in the research field, and for informing the practice of the design and support of collaborative activities. A major objective regards the development and use of analysis tools that are simultaneously appropriate for conducting meaningful evaluations of collaborative processes and efficient for practical use. This article adopts an innovative approach to CSCL analysis, involving the use of a rating scheme for the assessment of collaboration quality in several of its core dimensions. It involves appropriately trained evaluators that assign ratings on significant aspects of collaboration. Based on statistical analyses of ratings of 228 collaborating dyads working synchronously on a computer science problem-solving task, it reports and interprets interesting findings concerning general trends of collaborative practice, associations between dimensions of collaboration quality, and the way they relate to the quality of the problem's solution.

Introduction

Analysis and evaluation of CSCL activities are valuable for both building new knowledge in the research field, and for informing the practice of the design and support of collaborative activities. Several methodological approaches of analysis of CSCL have been suggested that range from deep-level qualitative analyses of small interaction-rich episodes of collaboration, to quantitative approaches focusing on measuring significant aspects of collaborative activities by statistical means (Stahl, Koschmann, & Suthers, 2006). Moreover, the goal of analysis is often shaped by practical purposes, aiming at regulating collaborative processes in real time (Soller, Martínez, Jermann, & Muehlenbrock, 2005), scaffolding learners in order to build new knowledge (Reiser, 2002), or informing the design and structuring of the conditions of the CSCL process (Kollar, Fischer, & Hesse, 2006), including the design of the tools that mediate them (Suthers & Hundhausen, 2003).

A major objective regards the development and use of analysis tools that are simultaneously appropriate for conducting meaningful evaluations of collaborative processes, rather than being based on "surface" counts of interaction events, and efficient for practical use. A rather newly applied approach in the field regards the use of *rating schemes* for evaluation of collaborative activities that involves human agents assigning ratings to predefined aspects of collaboration. In more detail, a rating scheme or a rating scale is "a measuring instrument that requires the rater or observer to assign the rated object to categories or continua that have numerals assigned to them" (Kerlinger & Lee, 2000, p. 736, cited in Meier 2005). Rating schemes are discriminated from coding schemes in that they are used to make a judgment on a larger piece of data each time, and are based on the knowledge and the critical skill of the human agent that applies them, in contrast to coding schemes that demand from the coder to neutralize the process by following strictly defined rubrics (Kerlinger & Lee 2000). So far, the rating approach has been applied in the CSCL field for assessing the "level of perspective taking" in asynchronous online discussions (Järvelä & Häkkinen, 2003) and the quality of collaboration through videoconferencing in synchronous problem-solving (Meier, Spada, & Rummel, 2007).

As an analysis tool, the rating scheme combines desirable properties of qualitative and quantitative techniques. Observed behavior is compared to a predefined standard that has been formed based on established CSCL theory and empirical analyses of typical collaborative sessions. This can then lead to *quantitative* judgments of the *quality* of collaboration. Therefore, the rating approach offers quantitative results that reflect subtle aspects of collaboration accessible to the human intellect but not easily detectable using any strict formalizations, or content analysis rubrics. Moreover, it is more time-efficient than deeper-level qualitative approaches, suitable for supporting CSCL practices, such as the provision of feedback to participants, targeted at the aspects of collaboration for which problems are reported. In addition, the rating approach allows assignment of grades by tutors for collaborative performance and not just for the correctness of the task's outcome. Ratings can also be useful for unraveling trends in collaborative practices in large populations, where qualitative approaches are not feasible, with the advantage that the ratings that constitute the object of analysis cover deeper-level aspects of collaboration than most other quantitative approaches. Finally, results of the rating process can be valuable as research aids, used as "quick indicators where more detailed analyses are merited, thereby focusing the detail work" (Stahl et al., 2006, p. 13).

This article builds on the rating tool developed by Meier et al. (2007), and adapted by Kahrimanis et al. (2009), and applies the approach to a large dataset of 228 synchronous collaborative problem-solving dyads. It investigates common trends in this population, related to the several dimensions of collaboration quality defined, and concludes to some interesting interpretations on factors influencing collaborative performance. In addition, the associations between dimensions of collaboration quality are investigated statistically, validating empirically the design of the rating tool. A third investigation made possible by the scale of the dataset relates collaboration quality with the quality of the outcome of the process, in this case the problem solution diagram. The intuitive assumption that good collaboration leads to good task performance is tested on statistical terms. The article concludes with discussion and proposals for further research built on this work.

Collaborative Setting

The study involved about 350 first year students of the department of Electrical and Computer Engineering of the University of Patras, Greece, engaged in jointly building the diagrammatic representation of an algorithm as an assignment that was part of an introductory to computing course. Randomly formed dyads of students interacted through Synergo (Avouris, Margaritis, & Komis, 2004), communicating via an integrated chat tool, and jointly designing a flow-chart representation of an algorithm on a shared workspace. Collaborative sessions took place in a university laboratory room and lasted from 45 to 75 minutes. Students were free to use their own resources such as textbooks, however the feedback they received to their questions was restricted to technical support. In order to motivate students to work collaboratively, they were informed that the grade would be formed 50% by the quality of their collaboration and 50% by the completeness and correctness of their joint solution. Moreover, students were given general instructions on what constitutes good collaborative practice according to the core dimensions of collaboration. Dyads were spaced in a way that partners communicated exclusively through Synergo.

The domain of the task was basic computer algorithms. Students were asked to build and investigate the correctness of elementary algorithms using flowchart diagrams (Bohl, 1971). The task given to students can be considered an “intellective task” with a "demonstrably correct solution" (Laughlin, 1980). The correctness of the solution is concretely defined, based on the notation used. All students were taught the knowledge demanded in order to handle the task sufficiently in university lectures before the lab sessions, even though some of them may have been already familiar with the task domain from secondary education.

The Rating Approach

Rating Scheme

The conceptual framework that shaped the analysis approach applied in this work was developed by Meier, et al. (2007). The framework was operationalized through a rating scheme. Due to significant differences between the setting that led to the definition of Meier’s, et al. (2007) scheme and the current setting, a laborious process of generalizing and adapting the initial framework to the current setting was followed (reported in Kahrimanis et al., 2009). The resultant scheme specifies seven core dimensions of collaboration quality, shown in Table 1.

Table 1: Dimensions of collaboration quality.

General aspect of collaboration	Dimension of collaboration	# Dim.
Communication	Collaboration flow	D1
	Sustaining mutual understanding	D2
Joint information processing	Knowledge exchange	D3
	Argumentation	D4
Coordination	Structuring the Problem Solving Process	D5
Interpersonal Relationship	Cooperative orientation	D6
Motivation	Individual task orientation	D7

Rating Procedure and Reliability

The rating procedure was carried out in two main phases. The first one, reported in detail in Kahrimanis et al. (2009), consisted of 101 dyads which were rated for each dimension by two raters with prior experience with the current setting after an extended pilot phase of training. The rating was conducted using video-like reproductions of the collaborative processes provided by the Synergo playback tool. The pilot phase involved the rating of the 1/3 of the dataset jointly by the two raters, aiming at overcoming discrepancies between them and potential misunderstandings on the rationale of the rating process, and simultaneously avoiding “over-training” that may lead to artificial consensus reached by succumbing to trivial aspects of collaboration. The second 1/3 of the dataset was rated in parallel by each rater individually and the comparison of concordance in their ratings led to very good inter-rater reliability results (Kahrimanis et al., 2009). The last 1/3 of the data was divided in two equal parts and each rater assigned ratings to each part separately.

The second phase, which was deemed necessary in order to extend the population of collaborative sessions, consisted of additional 149 dyads for which the setting of the labs was identical with the first phase, varying only in minor aspects of task details, being thus appropriate for integrated analysis. This way, large-scale statistical elaborations became possible. In the second phase, the ratings were applied by the same raters as in the first one and new tests of inter-rater reliability were made. The two raters co-rated 1/3 of the dataset (50) dyads, and split the remaining 99 dyads in two approximately equal parts, so that each rated half of them.

Overall results of inter-rater reliability of the whole dataset for each dimension of the rating scheme are shown in Table 2. For dimension D7, the reliability scores for the average rating (D7a) and the absolute difference between the ratings of the two students (D7b) are shown. The scores are good for all dimensions and indicate that the rating process was reliable according to established empirical rules (Fleiss, 1981; Cicchetti & Sparrow, 1981; Garson, 2009).

The final integrated dataset consists of the sum of dyads from the two phases (101 + 149 = 260) minus instances of collaborative sessions that were not considered due to technical problems in the logging mechanism of the Synergo tool caused by network failures during a few sessions.

Table 2: Reliability scores for each dimension of collaboration quality.

General aspect of collaboration covered	Dimension of collaboration	ICC	ICC (cons. adj.) = r	Cronbach's. α	Spearman's. ρ	Kendall's τ	Same rating
Communication	D1	.85	.86	.92	.85	.77	67%
	D2	.86	.87	.93	.84	.77	64%
Joint information processing	D3	.91	.91	.95	.89	.83	74%
	D4	.87	.87	.93	.86	.78	64%
Coordination	D5	.86	.86	.92	.83	.78	72%
Interpersonal Relationship	D6	.92	.92	.96	.89	.83	77%
Motivation	D7a (average)	.83	.84	.91	.80	.76	73%
	D7b (abs. diff.)	.88	.89	.94	.78	.74	74%
	Average D1-D6	.95	.95	.98	.93	.82	-

Measuring Dimensions of Collaboration Quality

Descriptive Findings

The first part of analysis regards descriptive statistical investigations of the dyads' collaboration quality for each dimension of the rating scheme. The current sample extends the size of a typical dataset used in CSCL research and can lead to interesting insights on distinct aspects of collaborative behavior as they emerge in real-world, large-scale activities. Therefore, descriptive measures such as the mean, median, mode, and standard deviation were calculated (Table 3), as well as the distribution of grades for each dimension (Figure 1).

Table 3: Descriptive metrics of collaborative dimensions of the rating scheme.

	D1	D2	D3	D4	D5	D6	D7a	D7b	Avg. D1-D6
mean	0.41	0.43	0.62	0.50	0.48	0.84	1.52	0.73	0.55
median	1	1	1	1	1	1	2	0	0.83
mode	1	1	2	2	2	2	2	0	2
std. dev.	1.24	1.31	1.34	1.44	1.35	1.33	0.67	1.15	1.14

The distribution of ratings for all collaboration dimensions indicates a positive bias, i.e. the dyads that were rated with a rating greater than 0 in all collaboration dimensions outnumber the dyads that were rated with a negative rating. Figure 1 illustrates the histograms of the distributions of values for dimensions D1 to D6.

The first two communicational dimensions (D1, D2) appear to have the lowest mean values compared to all others, although they are still positive. Dyads tended to score higher in information processing dimensions (D3, D4) than in communicational ones. Between these two dimensions, better ratings were obtained for D3, *knowledge exchange* (which describes lower-level practices of the information processing aspect), while D4, *argumentation* ratings followed a more diverse distribution. Ratings for D5 were more moderate, influenced by

45.6% of the dyads that where characterized by negative or neutral collaboration quality in *structuring the problem solving process*. On the contrary, D6, *cooperative orientation* was deemed successful for most of the dyads (65.8% got at least 1) and negative instances were rather rare, something reflected in the mean value of this dimension which is the highest reported among D1-D6.

Concerning dimension D7, which is used to assess *individual task orientation* at the level of each collaborating partner, its average scores (D7a) tended to be significantly higher than ratings of any other dimension. 56.1% of the dyads were assigned with the highest score in D7 for both the collaborating partners, while 71.5% had an average rating of at least 1. Cases where the average of the ratings for the two students was 0 or less were limited to 8.3% of the whole population. Concerning D7b, in most cases (63.2%) it was equal to zero, i.e. both students were assigned the same rating in D7. From this subset, 88.9% relates to cases where both students where highly committed to the task, while the remaining dyads were not assigned with any negative ratings either. Therefore, no dyads in the whole dataset were reported for which both of the partners were not motivated to solve the task. This can be explained by the circumstances of the lab, where at least one of the students took the responsibility to, at least, solve the problem on her own. On the other hand, in one out of five dyads (19.7%), significant imbalance in the commitment towards the task was noticed (difference of at least 2 grades in the scale), indicating that one student tended to take responsibility and dominate the process. Finally, the overall average rating of collaboration quality had a mean value equal to 0.55 and a median value of 0.83. 35.5% of the dyads had an average of 0 or less, whereas for the rest 64.5% the average was positive. 42.5% of the dyads had an average value of 1 or more.

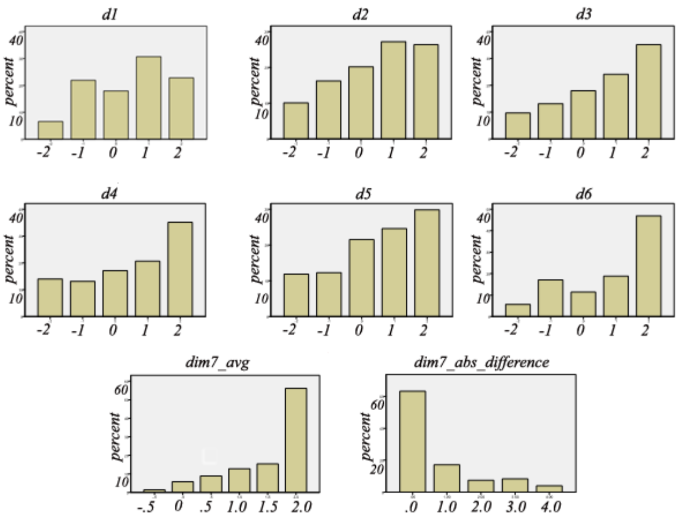


Figure 1. Distributions of values of dimensions D1-D6 of the rating scheme and the average and absolute difference in ratings of D7 (dim7_avg, dim7_abs_dif) for each participant.

Discussion on Measuring Collaboration Quality

On a general level, collaboration quality of the examined population of dyads was deemed successful in statistical terms. The relatively good performance of most dyads can be attributed to two main factors. First, a major influence was due to the motivations given to the students as 50% of their final grade was based on the assessment of the quality of collaboration; second, in the introduction of the collaborative activities by the tutors, students were encouraged to follow desired practices for each collaborative dimension.

Regarding the distinct dimensions of collaboration, students performed best at motivational and interpersonal aspects, they were mostly good at joint information processing aspects and had moderate performance on communicational and task coordination aspects. From these findings, a clear trend can be discerned that discriminates dimensions of collaboration quality between those that can be mastered by students based on their general or intuitive skills and those for which good collaboration quality can be achieved after a period of appropriation or if collaborative competences have been developed from prior experience. For the first category of dimensions, conscious intentions of collaborating partners to improve on them are usually achieved based on motivations and simple explanations of good practices in these specific dimensions, such as to communicate their task-related elementary knowledge, or to be pleasant and kind at the social level. On the other hand, general instructions given to students on how they should perform in dimensions such as *sustaining mutual understanding* and *structuring the problem solving process* are not so easily comprehensible by them. Moreover, even if instructions are communicated sufficiently, in order to achieve good performance more collaboration-specific skills are needed.

From this discussion, it can be inferred that, apart from giving sufficient motivations to students, a systematic approach towards achieving better collaboration quality in future activities should involve appropriate instructions on collaboration with more emphasis on the dimensions more difficult to learn. This can be done by arranging some targeted training sessions, by scripting the collaborative process in order to scaffold collaborative performance especially with regards to D5 (*Coordination*), or by providing feedback to participants if possible.

Associations between Dimensions of Collaboration

Associations between dimensions of collaboration quality can be conjectured by the definition of the rating scheme. Still, the extended dataset gathered offers the opportunity to validate such top-down assumptions empirically, by applying suitable statistical manipulations, as the relations between dimensions can be detected based on the ratings of collaboration quality applied in 228 cases. A suitable statistical technique is *MultiDimensional Scaling (MDS)* analysis based on the bivariate correlations between dimensions (Young, & Hamer, 1994).

In the specific case of this study, the unit of analysis of the technique is the collaborative dimension as it is defined by the rating scheme. Calculations are applied to a correlation matrix between the dimensions, as it was formed based on the ratings of each of the 228 dyads. The technique provides insightful two-dimensional diagrams representing the position of collaborative dimensions in such a way that dimensions correlated tightly are placed closer to each other in space than dimensions that do not relate that much. For the current application of the technique, disparities between correlations are represented with spatial Euclidian distances. The MDS algorithm used was SMACOF (Scaling by MAjorizing a COnvex Function) (De Leeuw, 1977).

Results and Internal Validation of the MDS Algorithm

The results of the application of the technique are depicted in Figure 2 (using Kendall's τ scores in the correlation matrix). A similar diagram was the result of the application of the same approach using Spearman's ρ coefficients instead (Kahrimanis, Chounta, & Avouris, 2010).

Moreover, in Figure 2, the Shepard diagram (Shepard, 1962) is included that serves for the evaluation of the algorithm (Steyvers, 2002). The filled circles in the diagram represent the Euclidean distances presented by the MDS algorithm, whereas the empty circles represent the distances calculated by the monotonic regression function of the algorithm (De Leuw, 1977). Moreover, Kruskal's stress for the application of the algorithm was measured at the acceptable level (Kruskal, 1964; Borg, & Groenen, 1997) of 0.062 for the concluding 28th iteration of the algorithm.

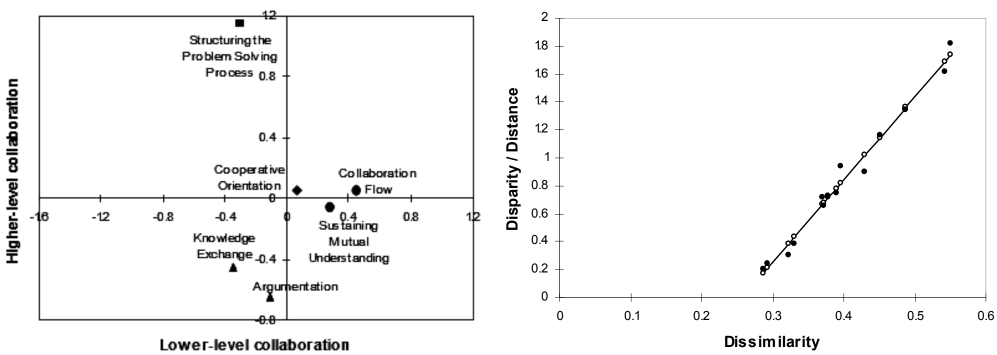


Figure 2. Multidimensional scaling diagram (left) and Shepard diagram (right) with collaboration quality dimensions D1-D6 using a similarity matrix based on Kendall's τ correlation score.

Interpretation of MDS Analysis

As it is evident from Figure 2, dimensions covering different aspects of collaboration quality cover different parts of the two-dimensional space. Dimensions covering the same aspect of collaboration (denoted by the same symbol in the diagram) stand close to each other. Regarding the interpretation of Figure 2, the coordinates of each dimension are used for the representation of its distance from other dimensions. Therefore, the range of each axis should be thought of as representing aspects that differentiate dimensions in the way they reflect different facets of collaboration. The rationale followed in diagram interpretation is described next.

Higher-order dimensions of collaboration are reported with higher absolute values on the vertical axis, while lower-level ones have higher absolute values on the horizontal axis (D6, *cooperative orientation*, which is placed near the zero point does not relate to any of these axes). Thereby, the vertical axis can be considered to stand for high-level collaboration aspects and the horizontal axis to stand for lower-level collaboration aspects. Concerning the horizontal axis, the two *communicational dimensions* (D1 and D2) are placed on the right of the diagram, taking positive values, whereas the two *information processing dimensions* (D3 and D4) are placed on the left, taking negative values. Thus, from left to right, the horizontal axis can be considered to designate the range from task-related low-level facets of collaborative activity to task-unrelated ones.

In the case of lower level collaborative activity, task-unrelated facets mostly refer to communication. *Collaboration flow* (D1) takes the largest positive (and absolute) value on the axis, since it constitutes the lowest-level dimension of the scheme. *Sustaining mutual understanding* (D2), on the other hand, is placed closer to zero and has a more noticeable Y coordinate. Among the information processing dimensions, *knowledge exchange* (D3) has the highest negative value since *argumentation* (D4) is related more to high-level collaborative activity. Coordination (*Structuring the problem solving process* D5) is also placed to the left of the Y axis. According to the proposed axes interpretation, this reflects the fact that D5 is shaped by task-related issues in the lower level of collaboration in contrast to task-unrelated aspects in higher-level collaboration. In general, results obtained from the MDS analysis are in accordance with the definition of the dimensions of the rating scheme. Distances represented by the algorithm are reasonable: a diagram of a similar rationale, applied by the researchers in a top-down manner, would probably resemble the one found empirically. Moreover, the approach offers subtler information on the exact associations between dimensions.

In conclusion, it should be noted that some properties of the definition of the axes are to some extent arbitrary and their relation to higher and lower-level aspects of collaboration constitute an interpretation rather than an "objective" result of the technique. For the output of the algorithm presented in Figure 2, the algorithm was initialized in such a way that the axes would be more interpretable, something that constitutes a common practice when applying MDS in other research domains (Guttman, 1968; Borg & Lingoes, 1987).

Relation between Collaboration Quality and Solution Quality

As already discussed, students were motivated for collaborating, while the outcome of the collaborative task, i.e. the algorithm diagram delivered, counted for 50% of the grade. The solution to the problem was graded separately, in a scale from 0 to 10. The assessments of the solution were quite straightforward since the task in all its variations had a demonstrably correct solution (Laughlin, 1980) with a few easily defined alternatives in parts of the diagrams, in some cases. The histogram of the grades (for N= 228 dyads) is provided in Figure 3.

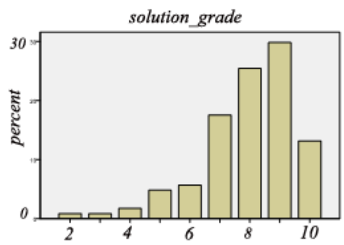


Figure 3. Frequencies of solution grades in the population in percentages.

The mean value for the sample was 7.96, with median = 8, and mode = 9 (68 times in 228 cases). 8.3% of the dyads were graded with 5 or less indicating a bad solution, 23.3% with 6 or 7 and 57% with 8 or more. Grades were generally high, something expected due to the nature of the tasks, which, although not too trivial, were of limited difficulty in order to allow researchers to focus their interest on the study of collaboration.

Table 4: Correlations between solution grades and dimensions of the rating scheme and their derivatives.

	D1	D2	D3	D4	D5	D6	D7a	D7b	Avg D1-D6
Solution	.248**	319**	.287**.	310**	.402**	.288**	.311**	-.238**	.349**
Grade	.301**	.391**	.352**	.380**	.487**	.344**	.369**	-.282**	.453**

** $p < 0.01$

The most interesting aspect related to the solution quality, regards the correlation of the solution grades with collaborative performance in general and with particular dimensions of the rating scheme in particular. Table 4 summarizes these correlations for the dataset (N=228). The values are Kendall's τ (the top cell) and Spearman's ρ (the bottom cell). For all combinations presented in Table 4, almost-medium to almost-high correlation scores between dimensions of collaboration quality and solution grade are observed. Still, the level of correlation differs significantly between different pairs.

The dimension that is correlated with the grade of the solution at the highest level is D5: *structuring the problem solving process*. This is reasonable because serious problems in the way dyads coordinate the collaborative process on a higher level, that often lead to mismanagement of their time, have often direct consequences on the completeness and correctness of the solution. The correlation is heavily influenced by cases where students do not manage to complete the task and therefore get lower grades. Next, D2: *sustaining mutual understanding* correlates at the second highest level with solution grade. It is thus apparent, that comprehensible communication between students can favor the development of a good solution, without, nevertheless, playing a determinant role. The correlation of the other communicational dimension: D1, *collaboration flow,* can be totally attributed to the variance it shares with D2. Moreover, information processing dimensions D3 and D4, are correlated at a medium level with solution grade, These are mostly related to the development and evaluation of the task per se and dyads that frequently exchange information and provide explanations to each other and get engaged in argumentation-intense collaborative sessions are more likely to develop a correct and complete diagram. D7: *individual task orientation* is also correlated with the solution grade judging by both its derivative measures. Summing up, with the exception of D2, in relation to the MDS analysis reported above, dimensions that are linked to higher-level aspects of collaboration are correlated with collaborative outcome at the highest extent.

Regarding the correlation between solution grade and average collaboration quality, which is $\tau=.349$, $\rho=.453$, it indicates medium to high correlation. This result is in accordance with the experience of CSCL practitioners that good collaboration is bound to relate to a good solution. It is also in accordance with similar evidence provided in some CSCL studies in different knowledge domains that prove the relationship between collaborative outcome and different kinds of measures of interaction (Chiu, 2004; Manlove, Lazonder, & de Jong, 2006). Nevertheless, the extent of this association is not so strong as to allow judgment of the solution quality by the measure of collaboration quality alone.

Conclusions and Further Work

This article presented the application of an innovative evaluation approach in a large set of synchronous collaborative activities. Based on an approach of adopting an existent rating scheme (Meier, et al. 2007) to the needs of the CSCL setting (Kahrimanis et al., 2009), the dataset used permitted various statistical elaborations.

Descriptive statistics revealed that, in relation to collaboration quality, students performed best at motivational and interpersonal aspects, were mostly good at joint information processing aspects and their performance lied at a moderately good level in communicational and task coordination aspects. It was thus indicated that the hardest aspects of collaboration are those that demand experience with the communicative setting and collaborative practice in general. Still, a majority of dyads got a positive grade on most collaborative dimensions. Success in collaboration was possible in most cases, based on general instructions from lab supervisors plus motivations for collaborating, through the grading rules used.

A multidimensional scaling analysis of the associations between dimensions of collaboration quality served to further validate the design of the rating scheme from an empirical standpoint, reassuring its initial assumptions and shedding light on the exact placing of each dimension in relation to others.

Solution quality was also correlated with dimensions of the rating scheme at a medium-to-high level. Task coordination and time management were the aspects correlating at the highest level with the correctness and completeness of the diagrams delivered, while, in general, dimensions covering higher-level aspects of collaboration were correlated more with solution quality than lower-level aspects. Findings reported are noteworthy, not allowing however assessment of collaboration quality to be based solely on solution quality.

Potential future work involves the use of the current approach and the findings obtained as a research aid for further deeper-level qualitative analysis of collaboration sessions of similar type. Moreover, the behavior of large populations of collaborative dyads in relation to dimensions of collaboration quality can be further investigated by applying other advanced statistical techniques such as clustering of collaborating dyads. A further goal regards the unraveling of common patterns of problematic dimensions of collaboration and trying to overcome them by appropriately restructuring the activity design. Furthermore, since, as it is claimed, the approach followed in this study leads to meaningful results in quantitative form, it can be used as a basis for evaluating less subtle logfile-based automated techniques of CSCL interaction analysis, and possibly reshape them so that they are optimally targeted towards reflecting substantial aspects of collaboration quality.

Endnotes

(1) According to an established empirical rule (Cohen, 1988; Hopkins, 2000), a correlation of $r = .1$ is considered to be low, of $r = .3$ medium, and of $r=.5$ or more high. This rule of thumb was initially proposed for Pearson product-moment coefficient r. These are implied here for Spearman's ρ scores (that resembles Person's r for non-parametric data).

References

Avouris N., Margaritis M., & Komis V. (2004). Modelling interaction during small-group synchronous

problem-solving activities: The Synergo approach, *ITS2004, Maceio, Brasil, September 2004*.
Bohl, M. (1971). *Flowcharting Techniques*. Chicago: Science Research Associates.
Borg, I., Lingoes, J. (1987). *Multidimensional Similarity Structure Analysis*. Beverley Hills: Springer-Verlag.
Borg, I, & Groenen P. (1997). *Modern Multidimensional Scaling*. Berlin: Springer-Verlag.
Chiu, C. H. (2004). Evaluating system-based strategies for managing conflict in collaborative concept mapping. *Journal of Computer Assisted Learning*, 20, 124-132.
Cicchetti D.V., & Sparrow, S.S (1981). Developing criteria for establishing the interrater reliability of specific items in a given inventory. *American Journal of Mental Deficiency*, 86, 127-137.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New Jersey: Lawrence Erlbaum.
De Leeuw, J. (1977), Applications of convex analysis to multidimensional scaling, in J. Barra, F. Brodeau, G. Romier & B. van Cutsem, (Eds), *Recent developments in statistics* (pp. 133-145). Amsterdam, The Netherlands: North Holland Publishing Company.
Fleiss, J.L. (1981) *Statistical Methods for Rates and Proportions*, 2nd. Edition. New York: Wiley.
Garson, G. D. (2009). Quantitative Research in Public Administration. Retrieved November 4, 2009, from http://faculty.chass.ncsu.edu/garson/PA765/pa765syl.htm.
Guttman, L. A. (1968). A general non-metric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika*, 33, 495–506.
Hopkins WG. (2000). A scale of magnitudes for effect statistics. In *A new view of statistics*. http://www.sportsci.org/resource/stats/effectmag.html#cohen. Retrieved on 21/08/2009
Järvelä, S., & Häkkinen, P. (2003). The levels of web-based discussions: Using perspective-taking theory as an analytical tool. In H. van Oostendorp (Ed.), *Cognition in a digital world* (pp. 77-95). Mahwah, NJ, USA: Lawrence Erlbaum Associates.
Kahrimanis, G., Meier, A., Chounta, I.A., Voyiatzaki, E., Spada, H., Rummel, N., & Avouris, N. (2009). Assessing collaboration quality in synchronous CSCL problem-solving activities: Adaptation and empirical evaluation of a rating scheme. *Lecture Notes in Computer Science*, 5794/2009, 267-272, Berlin: Springer-Verlag.
Kahrimanis, Chounta, I.A., Avouris, N. (2010). Determining relations between core dimensions of collaboration quality: A multidimensional scaling approach. Int. Conf. on Intelligent Networking and Collaborative Systems, INCOS, Thessaloniki, November 2010.
Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research*. New York: Harcourt College Publ.
Kollar, I., Fischer, F., & Hesse, F.W. (2006). Collaboration scripts–a conceptual analysis. *Educational Psychology Review*, 18, 159-182.
Kruskal, J. B. 1964. Multidimensional scaling by optimizing goodness-of-fit to a non-metric hypothesis. *Psychometrica*, 29, 1-27.
Laughlin, P. R. (1980). Social combination processes of cooperative, problem-solving groups on verbal intellective tasks. In M. Fishbein (Ed.), *Progress in social psychology, vol.1* (pp. 127-155). Hillsdale, NJ, USA: Lawrence Erlbaum.
Manlove, S., Lazonder, A. W., & de Jong, T. (2006). Regulative support for collaborative scientific inquiry learning. *Journal of Computer Assisted Learning*, 22, 87-98.
Meier, A. (2005). Evaluating Collaboration: A Rating Scheme for Assessing the Quality of Net-Based, Interdisciplinary, Collaborative Problem Solving. Unpublished diploma thesis, Institut für Psychologie, Albert-Ludwigs-Universität Freiburg im Breisgau, Freiburg, Germany.
Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *Int. Journal of Computer-Supported Collaborative Learning*, 2, 63–86.
Reiser, B.J. (2002). Why scaffolding should sometimes make tasks more difficult for learners. In *Proceedings of the Computer Support for Collaborative Learning (CSCL 2002), Boulder, Colorado, USA*, 255 - 264.
Shepard, R.N. (1962). Analysis of proximities: Multidimensional scaling with an unknown distance function I & II. *Psychometrika*, 27, 125-140, 219-246.
Soller, A., Martínez, A., Jermann, P., & Muehlenbrock, M. (2005). From mirroring to guiding: A review of the state of the art technology for supporting collaborative learning. *Int. J. on Artificial Intelligence in Education*, 15, 261–290.
Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences,* 409-426. Cambridge: Cambridge University Press.
Steyvers, M. (2002). Multidimensional Scaling. In *Encyclopedia of Cognitive Science*. London, Macmillan.
Suthers, D. D., & Hundhausen, C. (2003). An experimental study of the effects of representational guidance on collaborative learning. *Journal of the Learning Sciences*, 12(2), 183–219.
Young, F.W., & Hamer, R.M. (1994). *Theory and applications of multidimensional scaling*. Hillsdale: Erlbaum.

Technological Affordances for Productive Multivocality in Analysis

Gregory Dyke, Carnegie Mellon University, USA, gregdyke@gmail.com
Kristine Lund, CNRS, University of Lyon, France, Kristine.Lund@univ-lyon2.fr
Heisawn Jeong, Hallym University, Korea, heis@hallym.ac.kr
Richard Medina, University of Hawaii, USA, rmedina@hawaii.edu
Daniel D. Suthers, University of Hawaii, USA, suthers@hawaii.edu
Jan van Aalst, University of Hong Kong, China, vanaalst@hkucc.hku.hk
Wenli Chen, Nanyang Technological University, Singapore, wenli.chen@nie.edu.sg
Chee-Kit Looi, Nanyang Technological University, Singapore, cheekit.looi@nie.edu.sg

Abstract: Productive multivocality in CSCL has been the focus of a series of workshops involving the comparison and contrasting of multiple analyses of the same datasets, with the goal of learning how different epistemologies and analysis methods of collaborative learning can complement each other and allow a more complete understanding to emerge. A prerequisite to such work is the technological ability to assist the comparison of different analyses. In this paper, we show how the Tatiana framework for manipulating analytic representations was used to compare three different analyses of a computer-mediated small group problem solving session. In particular, we draw conclusions as to the technological affordances that are needed to ensure productive multivocality and illustrate the immediate benefits provided by the Tatiana framework.

Introduction

Researchers, designers and practitioners in the Computer Supported Collaborative Learning community work within a variety of epistemological frameworks, have multiple theoretical perspectives and use a variety of tools to apply different analytical methodologies to complex datasets. Such multivocality is a strength only to the extent that boundary objects can be created and shared in order to support dialogue between disciplines and approaches. Through a series of five workshops over four years, we have attempted to address the issue of what we call productive multivocality, particularly in the analysis of small group interaction in learning situations. Over the series of workshops (Suthers et al., 2011), we have come to realize that discussions surrounding these issues need to be grounded on multiple analyses of the same dataset and that, moreover, few differences can be addressed without diving back to the primary data and examining each analyst's approach.

Such comparison of analyses is hard to perform without adequate tool support and extends the need, already highlighted by many authors, for tool-support for analysis (Sanderson & Fischer, 1994; Hilbert & Redmiles, 2000; Suthers & Medina, 2008). However, most tools remain limited to certain kinds of media or certain kinds of analyses (e.g, Greenhalgh, French, Humble, & Tennent, 2007; Fiotakis, Fidas, & Avouris, 2007; Georgeon, Mille, & Bellet, 2006). Our ability to combine analyses through interoperating tools is further contingent on the design of models of analysis (Harrer et al., 2007). This ability is not only desirable within the context of productive multivocality but is necessary if researchers are to be able to analyse as a team (Goodman, Drury, Gaimari, Kurnland, & Zarella, 2006) and in order to validate and replicate existing studies (Reffay, Chanier, Noras, & Betbeder, 2008).

Suthers, Dwyer, Medina, and Vatrapu (2010) describe a generic analytic framework for computer-mediated interactions. They explain that such frameworks can be broken up into three foundations: an empirical foundation describes the real-world observable objects the data is composed of (typically events and relationships between events); a representational foundation describes how the empirical foundation is modeled and subsequently visualized (e.g. as a table, as a graph, etc.); a conceptual foundation describes the mapping of features of the other two foundations onto epistemological concepts (e.g. ideas and uptake). The Tatiana framework (Dyke, Lund, & Girardot, 2009) provides an extensible data model and a typology of operations from which a variety of representational foundations can be generated, resulting in the Tatiana tool (based on this framework) being generic enough to afford the creation and combination of a variety of analytic representations.

In this paper, we describe three analyses on the same dataset to illustrate how the Tatiana framework can be used to create and combine different analytic representations, and to show technological affordances that must be catered for when designing analytic tools. The complexity of this dataset and the diversity of analytical methodologies and representations that were used provide the basis for a discussion about the empirical and representational foundations of these analyses. Discussion of the conceptual or epistemological foundations and the commensurability and compatibility of various analytic approaches would need to be based on the entirety of datasets analyzed over the series of workshops and is beyond the scope of this paper.

In what follows, we first present Tatiana and the core concepts onto which data and analyses are mapped within the tool. We then describe the study from which the data was collected and the form in which it was shared among analysts. Three different analyses were carried out independently from each other with an overarching goal of identifying pivotal moments for learning. For both the data and the three subsequent analyses, we describe the initial analytic representations, their affordances, and the work that was performed, both to map them onto the Tatiana framework and to coordinate them in order to compare them. Finally, we show some examples of the immediate benefits of the Tatiana tool, suggest how it should be improved, and summarize the technological and pragmatic affordances that appear to be a prerequisite to multivocal analysis.

Tatiana: a Generic Analysis Environment

Tatiana (Trace Analysis Tool for Interaction ANAlysts) (Dyke et al., 2009) is an environment (and an underlying conceptual framework) designed for manipulating various kinds of analytic representations, in particular those that present a view on event-based data, be it the original data or subsequent analyses thereof. We call these representations *replayables*, because they can be replayed in a similar fashion to a video. Tatiana replayables comprise a sequence of events and benefit from Tatiana's four core functionalities, whose extensibility differentiates it from other tools: transformation, enrichment, visualization and synchronization (see Figure 1).

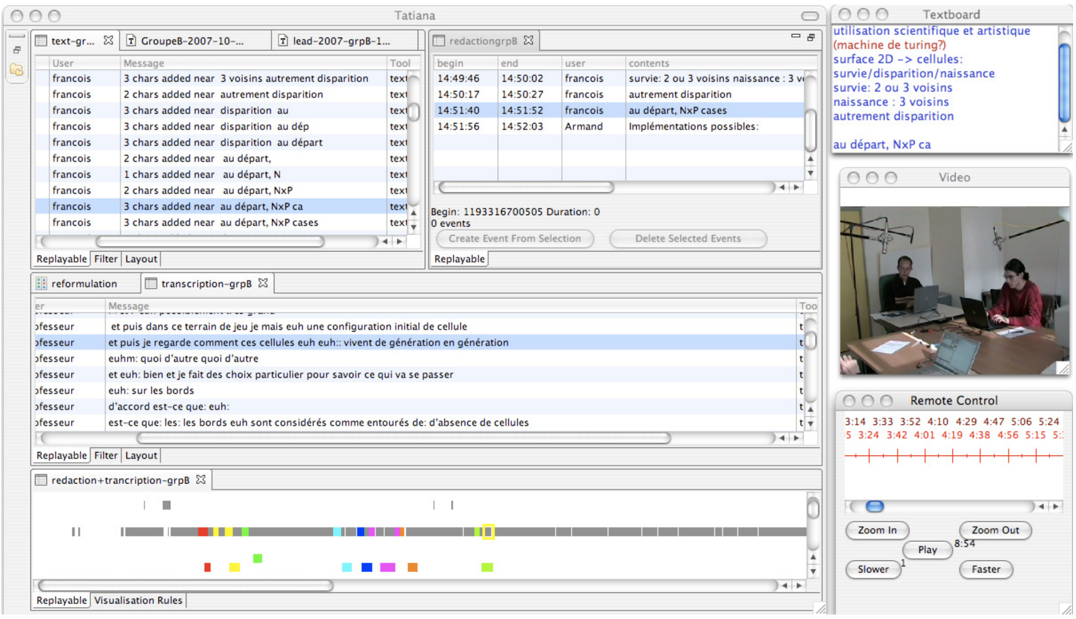


Figure 1. Various replayables visualized in Tatiana: traces of a shared text editor (top left), transcription (middle left), writing units (top center), visualization of reformulation (bottom left), synchronized with external tools, DREW replayer (top right), video player (middle right), remote control (bottom right).

Replayables can be transformed (automatically or manually) to create new replayables whose events might be a subset, an abstraction or a combination of events from other replayables. In analogy to a spreadsheet view of data, transformation can be thought of as creating, copying and deleting rows or events. In the remainder of this paper, we shall refer to events created as a combination or grouping of other events as *parent events*, since they create a parent-child relationship.

Events within Tatiana can be enriched by analytic data generated by the researcher. There are currently three kinds of enrichments supported by Tatiana: free-form annotation, categorization, and graphs. Categorization is simply a way annotating the events from a restricted list of words and can be used for coding, labeling and adding keywords. Graphs allow researchers to explicitly mark relationships between events. Conceptually, enrichments add additional columns (or properties) to events, or add links between events. Tatiana provides an extension mechanism to create new kinds of enrichment.

All replayables within Tatiana can be visualized in different viewers. Two viewers have been implemented so far: a table view in which data is presented as in a spreadsheet, with one row per event and columns for each of the event's properties; and a graphical timeline, which presents each event as a graphical object whose graphical properties (color, shape, size, position, etc.) can be set according to the properties of the event (user, tool, timestamp, category, etc.). Tatiana is extensible, allowing new kinds of viewers to be created, affording new ways of visualizing data.

Finally, all visualizations of replayables in Tatiana can be synchronized with each other and also with data viewed in external replayers such as media players. Synchronized replay means that, for example, during analysis of a video and its transcription in Tatiana, if a researcher clicks on a timestamped utterance in the table view, this action causes the replayer to move the video to this point. Conversely, when the video is replayed, the corresponding events in the transcript are highlighted one after the other. This functionality helps link different replayables together, allowing data to be looked at simultaneously from different angles.

Small Group Discussion about Fractions in Group Scribbles

The dataset we are reporting on was drawn from a three-year school-based research project using Group Scribbles (GS; Roschelle et al., 2007), which provides a shared representational space to support collaborative practices. Two fifth grade classes (about age 11) in a primary school in Singapore participated in the project. Students were asked to work in groups of four to solve the problem of dividing two pizzas equally among three children (Looi & Chen, 2010). They were first asked to work individually, either thinking about the solution or creating their private note, and then share it with the rest of the group.

Each student had an individual tablet PC with GS, which enabled them to create notes with hand-drawn sketches or typed text in their private space and then publish them to the group space. For this activity, the group space was prepared with a picture of two pizzas. Students could draw directly on the pizza (but were encouraged not to) or post notes in the public space. They could also view others' work, attach comments on other members' contributions, or move posts back to the private space.

The shared dataset focused on a group comprising four students whose names have been anonymized as Terry, Helen, Victor and Quentin. The students sat in pairs over two rows. One video camera was set to record the classroom session and another the target group. Each student also had a camera set up on the desk to capture their face. Screen capture software was used to record the contents of students' activities in GS environment. The resulting dataset comprises six videos (one of the classroom, one of the group, and one of each screen with the face video inserted into the bottom right corner). The episode that was analyzed lasted about 16 minutes.

During the episode, Victor and Helen both produce and discuss similar graphical solutions; Terry is initially confused and ends up producing a written solution that is equivalent to the graphical solutions. It is not clear to what extent Terry was unable to come up with any kind of graphical solution and to what extent his solution was influenced by Helen and Victor's solutions and by discussion with Helen. Quentin produces a symbolic arithmetic solution that is initially incorrect but gets amended later on as Quentin realizes his mistake.

Data Transcription and Subsequent Sharing within Tatiana

Transcripts of the videos were provided by Looi, Chen and Tan in a spreadsheet, with approximate time stamps (provided at one minute intervals) and with the activity organised into four columns, one for each student. Simultaneous actions (or near-simultaneous actions for which timing was not considered important by the transcriber) were transcribed in the same row, with actions taking place over a longer period of time spanning several rows. When the ordering of events was important to show potential reply-structures, these events were placed on consecutive rows. Many cells also contained screenshots to help understand the GS activity.

This kind of analytic representation is particularly interesting. It is able to condense the transcription into a reasonably small area (compared to a transcription with one row per utterance). It also accounts for the timing between events, albeit with a certain bias (as the transcriber decides whether precise information about how various actions overlap should be transcribed or not). This account can also be achieved at the relatively cheap cost of not having to find the precise beginning and end date of each event. Furthermore, the presence of screenshots helps to remove the difficulty in understanding the transcription without the aid of the video, making the transcription a remarkably standalone analytic representation. The cost for this is higher, as screen captures must be made and inserted for each event.

One goal of transcription is to remove the necessity of constantly having to replay the video. This comes at the cost of "trusting" that the transcription adequately serves as a proxy for the original video. In order to allow the analysts to view the transcription in synchronization with the videos and thus to alleviate this problem, the transcription was adapted to Tatiana before sharing. This produced four replayables, one for each student and a global replayable created through the automated merging of the four individual replayables. The events in these replayables were time-aligned with the videos, to reduce ambiguity and because Tatiana does not afford partial time-alignment. A basic typology of events was created (speech, note creation, posting a note, etc.) to enable better differentiation between them. Although no visualization in Tatiana allows screenshots to be associated with events, and the tabular view matches each event to a single row, the affordances present in the original transcription can be reproduced in Tatiana by opening in parallel the four replayables and one of the videos. However, as each video only shows the corresponding student's private space, all four videos must be opened to examine all the private spaces. The disadvantage of this mapping is that synchronization is dynamic, and static representations lose these affordances (i.e. it is no longer possible to click on a visualization in a paper and cause something to happen). One workaround is to print out the tabular view of the combined replayable, which at least provides complete timing information, in spite of displaying it in a less intuitive fashion.

Once this mapping into Tatiana was completed, the data was shared, both in Tatiana and in its original form. We now present the three analyses, two of which were performed with Tatiana and a third later adapted.

Analysis from the Perspective of Knowledge Building

Knowledge Building is the work that a community does to advance the state of knowledge in the community, such as creating, improving and using new ideas (Scardamalia, 2002). Knowledge building involves not only sharing ideas with others, but also jointly understanding and improving them. There must be a context in which such work is important, so that the new ideas are gradually used in the later work that the community does. Recent literature on knowledge building emphasizes "knowledge practices" (van Aalst, 2009; Hakkarainen, 2009). The first analysis was conducted from this perspective, entirely within Tatiana.

Because of the short length of the dataset, we did not expect evidence of all aspects of knowledge building – particularly the diffusion of the new ideas in the work of the community (a small group in this case). From the knowledge-building perspective, it is difficult to say whether learning has taken place without seeing ideas reused in multiple contexts, and by multiple students. Nevertheless it is possible to examine whether the efforts of students are limited to sharing ideas, or whether students collaboratively work to improve them and develop shared understanding. This analysis highlighted the fact that there were very few attempts to improve on existing solutions or to relate them (as all correct solutions are in some way equivalent). The problem itself was also not ideal, as it was not motivated by students' knowledge needs and was not very complicated: two of the students quickly identified solutions but did not understand why they should attempt to reconcile them.

Analytic Representations Created

The transcription was coded in Tatiana in a set of codes derived from the knowledge-building principles (Scardamalia, 2002) and some additional codes from a study that focuses on the social aspects of knowledge building (van Aalst, 2009). These codes were not mutually exclusive and each line was coded with all applicable codes. Tatiana's current implementation does not allow for more than one code per event (which might be described as *tagging*), so three enrichments based on the same coding scheme were created and most events were given between one and three codes. A further annotation enrichment was created to provide a narrative description of what was happening at each line and to justify the choice of certain codes.

Analysis of the Uptake of Representational Practices

Medina conducted uptake analysis of the GS dataset using concepts and methods from the uptake analysis framework described in (Suthers, Dwyer, Medina, & Vatrapu, 2010). This analysis was mostly performed within Tatiana. Its focus was twofold. The first part of the analysis showed, through participants' orientation to each other throughout the first few minutes that, in spite of little apparent explicit communication, the group was aware of the GS workspace as a shared medium for whose content they were accountable. In this brief period they also demonstrated an appropriation of a diverse range of modal resources (talk, gesture, and inscription), The second part of the analysis used uptake graphs and information about the context in which the dataset was produced. It showed that (1) the students were operating on representational practices developed in prior classroom activity (2), each student demonstrated a varied interpretation of these practices in the problem setting and (3), interaction between two of the students (Helen and Terry) revealed that, at times, reconciliation of interpretations can be mediated through shared, yet uniquely applied, representational and inscriptional practices (Medina, Suthers, & Vatrapu, 2009).

Analytic Representations Created

Similar to other analyses conducted under this framework, (e.g. Medina et al., 2009), the analytic implementation in the current investigation was multilevel. At an empirical level, the analysis attends to the sequential order of participants' actions in order to identify contingencies between acts and produce a contingency graph that, in turn, serves as a resource for identifying uptake. Identification of uptake relations, based on the empirical evidence provided by contingencies, produces an uptake graph as a derivative of the contingency graph. In the process of creating these graphs, a new replayable was created with its events being either a single act, already present in the original transcription, or a new event, representing a collection of related acts in the original transcript (e.g. when an idea was published in two separate notes, forming a single whole). This replayable added the affordance of being able to see how much time was taken between beginning the crafting of a new idea and its final publication.

Two graph enrichments were created, one for each analysis, making explicit the contingencies between the different acts. The replayable can be visualized on a graphical timeline, with one horizontal line per student. The arrows provided by the graph enrichment show how students re-use and respond to each other's contributions. An additional annotation enrichment was created to provide a narrative of the episode.

Analysis from a Cognitive Perspective of Group Understanding

The last analysis was performed in a spreadsheet and later adapted to Tatiana. The analysis was an attempt to further our understanding of how groups operate as a cognitive unit in artifact-mediated discourse situations (Jeong, Chen, & Looi, 2011). Jeong examined how group understanding developed by tracing contributions to the group workspace and how individual contributions evolved over verbal and non-verbal activities. This was done by splitting the session up into various activity threads around contributions that represent a group of postings about the same core idea. The contributions were further coded in terms of whether they were on-task or off-task, collaborative or individualistic, and made public or kept private.

The main findings were that the group space became crowded with many postings and comments, but still remained fragmented at the end of the activity, as students made little attempt to integrate and consolidate different postings and their relationships to each other. The analyst suggests that this could be remedied through better regulation of the group space and of the student activity. Furthermore, the students' group work was mediated by both verbal and non-verbal exchanges, e.g. a verbal question was answered by an action in GS. Thus, when group work is mediated by artifact construction, non-verbal actions must also be taken into account.

Analytic Representations Created

As with the original transcription, this analysis was particularly interesting from a representational standpoint. The original events were taken and rearranged, so that each event (or rather the cell in the spreadsheet representing this event) was placed in the row corresponding to the contribution thread it belonged to. This extracted the time-student ordering of the data into as many timelines as there were threads, thus enabled the analyst to examine more clearly how each thread developed, how many participants were involved and to distinguish off-task threads and exclusively private events more easily.

This representation can be mapped onto the Tatiana framework in any one of three ways, depending on the desired affordances. First, a replayable can be created in which the events of each contribution thread are collected together into a single parent event. In the Tatiana framework, this is a transformation that abstracts up from *individual events* to higher-level *threads*. Of the three, it is the most visually similar to the original and, via synchronization, allows events to stay in their original context while still showing to which thread they belong. Second, a categorization enrichment could enable the coding of each event according to the parent thread to which it belongs, this makes it easier to visualize thread-belonging as an extra property of the data, rather than as its central property. Third a replayable could be created for each thread, making it easier to consider threads in isolation. In the work done for this paper, only the first (initially) and second (later) was created.

Combining Analyses in Tatiana

The original idea was that the analyses (four enrichments for the first analysis, two enrichments and a replayable for the second and one replayable for the third) could be opened at the same time and visualized in synchronization in order to compare them, thus creating a common boundary object. In practice, this was not immediately possible, for several reasons.

First, the knowledge building categories were applied to all events. GS forces users to first write a note in their private space (producing a first event in the transcription) and then drag it to the public space, (producing a second). The first event would typically be coded "new idea" and the latter "community knowledge". The uptake analysis only considered events in the public space. In order to meaningfully compare the analyses, the knowledge building categories from events in the private space needed to be transferred to their public space counterparts.

Second, the replayable to which the uptake graph was applied contained events that were parents of the events in the original transcription (typically, the events of note creation, editing, publication and verbal explanation would all be grouped together into a single parent event). As enrichments (categories, annotations and links) attach to individual events, the newly created parent event was not aware of the KB categories and the children of the parent event were not aware of the links that had been attached to their parent event. Again, there needed to be some kind of transfer of enrichments for comparison to be possible.

Last, the initial replayable that was created for the group understanding analysis, with one event per thread, also did not afford easy comparison with the other analyses, since it was only possible to see which events belonged to which thread through synchronization.

To solve these problems, all three replayables (the original transcription, the one on which the uptake graph was drawn, and the threads replayable) were opened side by side along with their associated enrichments and synchronized. From these, a fourth replayable was created, which initially contained all the events of the transcription. The following procedure was then applied, event by event:

- The event was deleted (from the fourth replayable, but remained present in the original transcript), unless one of the following conditions was met:
  - It appeared (itself or by proxy via its parent event) in the uptake graph.
  - It was public, and coded otherwise than "off task" in the knowledge building analysis.
  - It was public and did not belong to an "off task" contribution thread.
- If the event was deleted and was private, its knowledge building codes and annotations were transferred to the corresponding publication event.
- If the event was not deleted, it was coded according to the thread it belonged to.

- If the event was the last publication event and child of a parent event in the uptake graph, it was added into a new graph in place of the parent event.

In order to preserve the affordance, provided by the uptake graph, of "seeing" how long a given idea takes to construct, we also then added back in the private events that lead up to a publication. The application of this process resulted in a single replayable containing only the events that had been considered pertinent in at least one analysis. Although "off task" activity, particularly in the regulation of social relations, has often been considered important in CSCL (Baker, Andriessen, & Lund, 2009), none of the three analyses focused on such events, which is why they were mostly ignored. In effect, the publication event of a note was considered as a proxy for its creation and editing on the one hand and for any parent events to which it belonged on the other, in order to conciliate the units considered in the various analyses.

In an attempt to provide a single visualization in which to examine all the analyses (in combination with synchronization with other visualizations), we used Tatiana to automatically produce a graphical timeline based on the various enrichments of the different analyses, an extract of which is shown in Figure 2. In this timeline, time progresses horizontally, each row corresponds to a contribution thread, each color corresponds to a student, squares are speech events, circles are GS-mediated events, small circles are private events, and the largest circles and squares show events which have been coded with the "core" principles of knowledge building (i.e. new ideas, idea diversity, etc.). The majority of the interaction occurs in the shown threads, with the core of the uptake between Helen (red) and Terry (blue). Potential for knowledge building can also be found in threads that do not involve multiple students (e.g. Quentin's symbolic solution in yellow).

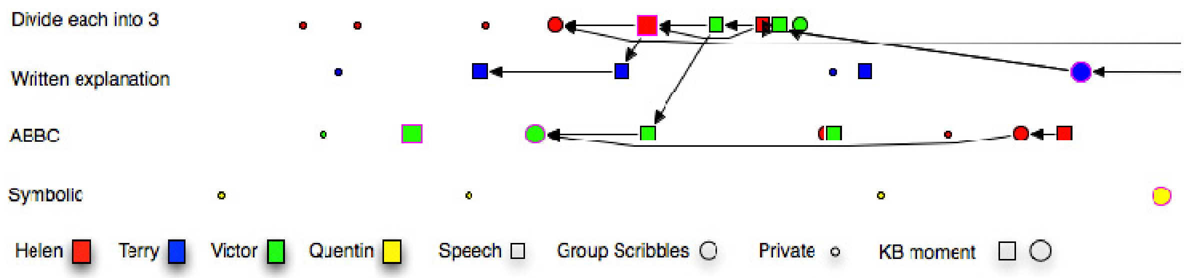


Figure 2. Graphical Timeline Showing the Combined Analyses by Thread. Four of the Nine Threads are Shown.

Benefits of Combined Analyses

Although we do not discuss the lessons learnt from these different analytic interpretations in this paper, this section briefly discusses some of the benefits that are immediately afforded by tools such as Tatiana when embarking on a multivocality or mixed-methods route.

The separation into threads carried the risk of having difficulty in identifying any transitions of ideas between threads. Opening the uptake graph on top of a graphical timeline where each line represents a thread, as in Figure 2, confirms that such transitions do occur. Conversely, the original uptake analysis showed that there was some uptake between Helen and Victor but focused on the importance uptake between Helen and Terry. However, this is not as obvious in the uptake analysis' original graph (cf. Figure 3) where the separation of users on different lines seems to indicate that the main interaction is between Helen and Victor. Separating by contribution thread, rather than by user, immediately focuses our attention on the places where uptake moves from one contribution thread to another, showing the interactions beween various students' solutions and representations. Indeed, the places where arrows cross from one thread to another in Figure 2 were identified by all three analysts as being of interest. This is hardly surprising and confirms the suggestion by Suthers and Medina (2008) that, in process analysis, trajectories of ideas should be examined to look for their transformation (which is an important concept in many epistemologies of collaborative learning and meaning-making).

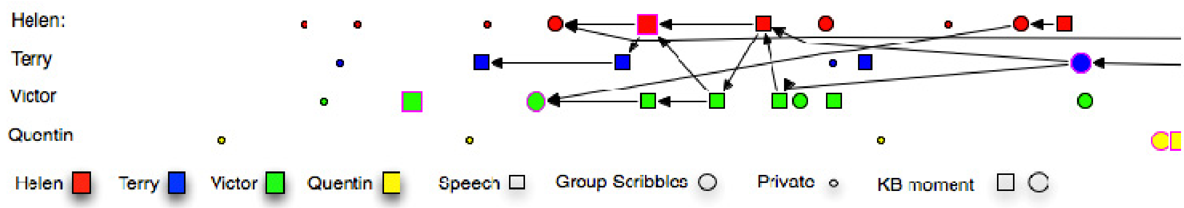


Figure 3. Same Data as in Figure 2 Above, But Organized by User, Rather Than by Thread.

This visualization also immediately highlights some differences between analytic approaches, providing the opportunity to discuss issues concerning the data (uptake between threads, important moments for knowledge building which are "ignored" by other analyses, etc.), which are connected to discussions about epistemological questions in general, grounded on these analyses, such as the relative importance of idea diversity (the mere "presence" of different ideas), collaboration in threads, or uptake across threads. Furthermore, when discussing data points, synchronization allows a return to the original data in the videos in order to verify that interpretations are not influenced by imprecision in the transcription.

Design of Tools to Support the Construction and Use of Analytic Representations

We now draw upon this case study to discuss how tools and their design can support multivocal analytic dialogue. While the previous sections show the Tatiana framework to be an important contribution as a means to reuse and combine analyses in order to create boundary objects for discussion of multiple analyses, we also highlighted some issues, which we hope to address and encourage other analysis tool designers to draw upon.

The biggest difficulty we had in combining these analyses stemmed from the unit of analysis. As explained by Suthers et al. (2010), the events of a CSCL transcript are frequently used as proxies for higher-level concepts. In other words, what is of interest to analysts are not so much events, but, for example, the ideas represented by these events. It is therefore not surprising that the proxy for an idea should be quite variable (initiation of private authorship, editing, publication, or some combination of the three). The issue can be addressed pragmatically, either by discussion after analysis (as exemplified in this paper), or by prior agreement among analysts. In either case, analysts might address what the unit of analysis should be, whether private spaces should be considered, whether the time taken to craft a contribution is important, and whether events in all modes (e.g. speech and tool mediation) should be considered equal.

A complement to this pragmatic approach would be to adjust tool support so that enrichment be transitively propagated to parent and child events under certain conditions. For example, it would make sense for the parent events in the uptake analysis to "inherit" the knowledge building codes attributed to their constituent child events. And in the case of the replayable containing contribution thread events, it would be beneficial to be able to propagate the thread an event belongs to down to the event itself in order to add the thread as a category. In general it should be easy to transition events across such near equivalent features as parent event belongingness, replayable belongingness and having identical properties (same user, same category, etc.).

During the course of this work, Tatiana's synchronization feature had to be refined. Sometimes synchronization is used to identify overlapping or synchronous events (e.g. between speech and artifact-mediated dialog); in other cases it can be used to identify the same event in different visualizations of the same replayable; a hybrid of these two is identifying the children of a parent event from one visualization to another. This only becomes a problem when there are many occurrences of simultaneous events, as is the case with the GS dataset. For this reason, Tatiana has been modified to make it possible to choose the synchronization mode: by timestamp or by identity (the latter includes parent-child event relationships).

The affordances synchronization provides are dynamic and lost in the case of static representations (such as those found in conference and journal publications). In spite of this drawback, there are so few static visual dimensions (size, shape, position and color) that it makes sense to transfer as many aspects as possible to a dynamic register, particularly for the *performance* of the analysis, as opposed to its publication.

Finally, so many analytic representations were created during the analyses and their combination that it would be useful to have meta-data describing their assumptions and purposes, along with software support for recording this meta-data. This is an extension of the recommendation by Reffay et al. (2008), that it is indispensable to include the educational and research context when sharing learning data.

Conclusion

In this paper, we described the lessons learned from the multivocal comparison of three independent analyses of the same dataset, and discussed their implications for the design of analytic tools. The original data set and the three analyses provided a diversity of analytic representations whose affordances were discussed. We then explained how these analyses were combined with the assistance of the Tatiana tool and provided some examples of the benefits of combining analyses in such a way. Finally, we discussed some of the issues raised and provided guidance for development of analytic tools beyond the existing Tatiana framework, focusing in particular on aspects such as analytic unit, propagation at different levels of abstraction, synchronization, and analytic and pedagogical context.

The current Tatiana framework proved a very positive experience as a means to create boundary objects for discussing multivocal analysis. These analyses provide further support for the hypothesis that Tatiana's provisions for generating representational foundations are generic enough to serve as the basis of boundary objects in CSCL. While there appears to be a certain influence of the conceptual foundation of an analysis on its empirical foundation (e.g. what aspect of the data will serve as proxy for an *idea*), this need not be an obstacle to multivocality, assuming analysts agree beforehand on a common empirical foundation (and its mapping onto respective representational and conceptual foundations), and are aware of the choices that might later introduce bias. Furthermore, the use of synchronization, automated transformation and enrichment reduces the need of finding a unique representational foundation for comparing analyses.

We hope to draw on this work to apply similar methods in comparing analyses in other datasets. In particular, we need to extend these methods to address analyses that are more quantitatively oriented. Having diminished the pragmatic representation-driven issues related to multivocality, we hope to be able to focus on methodological and epistemological issues to promote better understanding in the diverse field of CSCL.

References

Baker, M., Andriessen, J., Lund, K. (2009). Socio-relational, affective and cognitive dimensions of CSCL interactions: integrating theoretical-methodological perspectives. *CSCL'09*. Rhodes, Greece.

Dyke, G., Lund, K.,& Girardot, J.-J. (2009). Tatiana: an environment to support the CSCL analysis process. *Proc. CSCL 2009*, Rhodes, Greece, 58–67.

Fiotakis, G., Fidas, C. & Avouris, N (2007). Comparative usability evaluation of web systems through ActivityLens, *Proc. PCI 2007*, Patras, Greece.

Georgeon, O, Mille, A, & Bellet, B.T. (2006). ABSTRACT: a tool and a methodology for analyzing human activity mediated by a complex technical artifact. *Ingéniérie des Connaissances. Semaine de la connaissance.* Nantes, France.

Goodman, B. A., Drury, J. Gaimari, R. D., Kurland, L., & Zarrella, J (2006) Applying User Models to Improve Team Decision Making. Retrieved April 10, 2008 from http://mitre.org/work/tech_papers/tech_papers_07/06_1351/

Greenhalgh, C., French, A., Humble, J. and Tennent, P. (2007) Engineering a replay application based on RDF and OWL, *Online Proceedings of the UK e-Science All Hands Meeting 2007*, September 10-13, Nottingham: NeSC/JISC.

Hilbert, & Redmiles. (2000). Extracting usability information from user interface events. *ACM Comput. Surv.*, *32*(4), 384–421. doi: http://doi.acm.org/10.1145/371578.371593.

Harrer, A. Zeini, S. Kahrimanis, G. Avouris, N. Marcos, J. A. Martinez-Mones, A. Meier, A. Rummel, & N. Spada, H (2007) Towards a Flexible Model for Computer-based Analysis and Visualisation of Collaborative Learning Activities, *Proc. CSCL 2007*, New Jersey, USA

Hakkarainen, K. (2009) A knowledge-practice perspective on technology-mediated learning. *ijcscl 4*(2), pp. 213-231.

Jeong, H., Chen, W., & Looi, C. K. (2011). Analysis of group understanding in artifact-mediated discourse. *CSCL 2011*. Hong Kong.

Looi, C.-K., & Chen, W. (2010). Community-based individual knowledge construction in the classroom: A process-oriented account. *Journal of Computer Assisted Learning*, 26(3), 202-213.

Medina, R., Suthers, D., & Vatrapu, R. (2009). Inscriptions becoming representations. *CSCL'09*. Rhodes, Greece.

Reffay, C., Chanier, T., Noras, M., & Betbeder, M.-L. (2008) Contribution à la structuration de corpus d'apprentissage pour un meilleur partage en recherche, *STICEF journal*, Volume 15, ISSN : 1764-7223. Put online 14th October 2008, http://sticef.org.

Roschelle, J., Tatar, D., Chaudhury, S. R., Dimitriadis, Y., Patton, C., & DiGiano, C. (2007). Ink, improvisation, and interactive engagement: Learning with tablets. Computer, 40(9), 38-44.

Sanderson P., & Fisher C. (1994). Exploratory sequential data analysis: foundations. *Hum.-Comput. Interact.*, *9*(4), 251–317.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.) *Liberal education in a knowledge society* (pp. 67-98). Chicago: Open Court.

Suthers, D. D., Lund, K., Dyke, G., Rosé, C., Law, N., Teplovs, C., et al. (2011). Towards productive multivocality in the analysis of collaborative learning. *CSCL 2011*, Hong Kong.

Suthers, D. D. & Medina, R. (2008). Tracing Interaction in Distributed Collaborative Learning . *Annual Meeting of the American Educational Research Association* (AERA), New York, March 24-28, 2008.

Suthers, D. D., Dwyer, N., Medina, R., & Vatrapu, R. (2010). A framework for conceptualizing, representing, and analyzing distributed interaction. *ijcscl 5*(1), pp. 5-42.

van Aalst, J. (2009) Distinguishing knowledge-sharing, knowledge-construction, and knowledge-creation discourses. *ijcscl 4*(3), pp. 259-287.

Acknowledgments

The authors would like to thank the organizers and financers of the series of workshops on productive multivocality, as well as the many participants whose contributions helped structure our thoughts on these issues.

The Automatic Assessment of Knowledge Integration Processes in Project Teams

Gahgene Gweon, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA, ggweon@cs.cmu.edu
Pulkit Agrawal, Indian Institute of Technology, Kanpur, Uttar Pradesh 208016, India, pulkit@iitk.ac.in
Mikesh Udani, Indian Institute of Technology, Kharagpur, West Bengal 721302, India, mikesh.iitkharagpur@gmail.com
Bhiksha Raj, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA, bhiksha@cs.cmu.edu
Carolyn Rose, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA, cprose@cs.cmu.edu

Abstract: Automatic assessment of group processes in collaborative groups is one of the holy grails of the computer supported collaborative learning community. The conversation in collaborative work provides an important window into the inner workings of a group. In this paper we present work towards detecting where students are displaying "reasoning" in conversational speech and how others are building upon those expressions of reasoning ("idea co-construction (ICC)"). Such technology would add to the body of work in educational data mining another means of monitoring student work as well as contributing to the area of automatic collaborative process analysis. We begin by discussing our operationalization of targeted group processes, namely reasoning and ICC. We then discuss the level of success we are able to achieve applying machine learning technology to replicate this human analysis using simple audio signal processing techniques.

Introduction

As communication technologies such as cell phones and voice over IP become more ubiquitous and allow for communication and collaboration over multiple modalities including video, audio, and text to be accessible any time and any place, the line between online group learning and face-to-face group learning begins to blur. Furthermore, as more and more collaboration takes place over video and audio channels, the need grows for the CSCL community to think about how to extend collaboration support technologies from the text realm into audio and eventually video. In this paper we present work towards assessment of group processes from speech data, specifically focusing on processes related to knowledge transfer and knowledge integration within groups. Specifically, we target design project classes, which present challenges both for supporting and for assessing learning because the learning is self-directed and knowledge is acquired as needed throughout the design process. What makes it especially tricky from an instructor perspective is that regardless of whether instructor supervision takes place online, in a whole class setting, or in face-to-face advising meetings with student groups, the bulk of student learning takes place without the instructor present. While this provides students with opportunities to develop skills related to "learning to learn", it can also mean that instructors are left not knowing when and how they can intervene to support the students most effectively. It is well known from the social psychology literature on group work that groups frequently do not function in an ideal way (e.g., Faidley et al., 2000).

Prior work investigating assessment practices of project course instructors reveals both the importance and difficulty of accurately assessing important group processes (Gweon et al., 2011). In this work, project course instructors reported attempting to assess groups in terms of planning and goal setting, productivity and progress, knowledge sharing and group knowledge integration, leadership and division of labor, and interpersonal dynamics. Because each student's expertise and experience provides an important added dimension to the project development process in a multi-disciplinary team, at the level of knowledge sharing and group knowledge integration, instructors wanted to see students behaving as intellectual leaders within their groups, taking the initiative to contribute their own unique expertise and perspective to the group. Beyond that, they wanted to see the contributed ideas taken up and transformed by the group as evidence that the end product would represent a true integration of expertise across the students within the group, and not just a patchwork product that frequently results from dysfunctional group efforts.

In our work we operationalize group knowledge integration processes through an adaption of the construct of transactivity from the field of collaborative learning (Berkowitz & Gibbs, 1983; Teasley, 1997; Weinberger & Fischer, 2006), which we refer to as Idea Co-Construction (ICC). Transactivity has its foundations in Piaget's theory of learning, and is theorized to provide opportunities for cognitive conflict to be triggered within group interactions, which may eventually result in cognitive restructuring (de Lisi & Golbeck, 1999). The construct of transactivity makes several important distinctions related to the notion of intellectual leadership and group knowledge integration that are relevant for our work. First, is the important distinction between contributions that visibly display reasoning behind an assertion versus contributions where the reasoning that lead to an assertion is hidden. Assertions that display reasoning are further subdivided into ones that build on or operate on prior assertions, which are the transactive variety, and ones that represent a new direction in the conversation, which are externalizations. Externalizations position students as intellectual leaders within a conversation. However, true leadership requires that the leader is received as such by the other group members. Thus, externalizations that are not followed by transactive contributions building on them may be regarded as failed attempts at intellectual leadership. A more complete picture of intellectual leadership within a group can be obtained through tracking the distinction between assertions that do not make reasoning visible, externalizations, and idea co-construction. Our notion of ICC is different from earlier characterizations of transactivity. For example, our notion of ICC does not necessarily require that expressions of reasoning involve a comment on or operation on articulated reasoning that was previously contributed. Instead, the expression of reasoning may simply integrate information articulated previously even information articulated in such a way that it does not conform to our operationalization of reasoning. We adopt this slightly more relaxed notion of "building" in order to be more inclusive of the types of integrative contributions that students contribute since in our experience, the ideal of transactivity is not frequently achieved.

In our work we build on recent efforts to support instructors in managing groups by offering them forms of automatic assessment and reporting. In prior work, researchers have looked at automatically detecting various aspects of student activities during their work together (Kay et al., 2006; Pianesi et al., 2008). Various forms of data have been used including message board postings (Kim et al., 2007), chat data (Soller & Lesgold, 2003), video (Chen, 2003), and audio (DiMicco, et al., 2004). Our contribution is technology to use speech data to distinguish between ICC, externalizations, and contributions that do not display reasoning. Thus, in the remainder of the paper we first situate our work in the midst of current directions in speech processing. Next we discuss our approach to operationalizing reasoning displays and ICC. We then move on to a discussion of the technological contribution of the paper. Finally, we present an evaluation of our approach and conclude with a discussion of current directions.

Motivation and Background

Automatic analysis of ICC and related constructs such as transactivity is not a completely new direction in the CSCL community in itself. However all of the prior published work was related to automatic processing of text, such as newsgroup style interactions (Rosé et al., 2008), chat data (Joshi & Rosé, 2007), and transcripts of whole group discussions (Ai, Kumar, Nguyen, Nagasunder, & Rosé, 2010). One lesson learned by comparing across efforts to detect transactivity in a variety of types of interactions is that a key feature enabling high accuracy of recognition is being able to measure content similarity between a contribution and the contributions from other conversational participants that occurred within the same topic segment earlier within the conversation. For example, Rosé and colleagues (2008) report that in a classification task with a coding scheme related to transactivity, adding a single feature representing content similarity with prior contributions within the same thread from other participants to a baseline feature space, keeping all other aspects of the modeling technique constant, produced an increase in agreement with human coding from 0.5 Kappa to 0.69 Kappa.

The unique contribution of the work presented here is that it is not applied to text, but to recorded speech. Although the speech data we work with has been transcribed prior to the annotation process, the automatic analysis technique we describe does not use the transcriptions as input. Rather, the speech signal is first processed using basic audio processing techniques in order to extract features from the segments of speech, which are then used for classification using a machine learning model. One might assume that the most straightforward approach would be to use speech recognition technology to transform a speech recording into an automatically obtained transcript and then simply apply a model such as the one developed by Ai and colleagues (2010), which was applied to transcriptions of face to face interactions. However, the state of the art in speech recognition is still too poor to make this a viable option. Although some tutorial dialogue systems such as Scot (Pon-Barry, et al 2006) and ITSPOKE (Forbed-Riley & Litman, 2009) have used speech recognition technology to detect uncertainty in student responses, neither systems required high accuracy of the content in order to make this attribution. For instance, both systems used speech recognition to detect lexical hedges (e.g. I think, I thought, maybe) or pauses to detect uncertainty in student responses. Therefore, despite the great potential value in automatic transactivity or ICC analysis directly from speech, considering the importance of content similarity with prior contributions evidenced in prior work, it remains to be seen what level of accuracy is possible just from the speech signal itself.

The technique we evaluate in this paper is related to prior work on speech processing for other classification tasks. There has been some prior work on automatic assessment of group interactions in the CSCL community focusing on speech as input (DiMicco et al., 2004; Gweon, Kumar, & Rosé, 2009), however that work was more focused on the amount of contribution from each speaker overall rather than anything specific related to the nature of individual contributions. In the language technologies community, some prior work has focused on the nature of conversational contributions, however. For example, Ranganath and colleagues (2009) used acoustic and prosodic features extracted from speech data to predict whether a speaker came across as flirting or not in a speed dating encounter. Similarly, Ang (2002) and Kumar and colleagues

(2006) applied a similar technique to the problem of detecting emotions such as boredom, confusion, or surprise, whereas Liscombe (2005) applied the technique to the problem of detecting student uncertainty. All of this work makes use of signal processing techniques that are able to extract basic acoustic and prosodic features such as variation and average levels of pitch, intensity of speech, amount of silence and duration of speech.

Acoustic and prosodic features are frequently associated with intuitive interpretations that make them an attractive choice to play a role in baseline techniques for these stylistic classification tasks. For example, increased variation in pitch might indicate that the speaker wants to deliver his ideas more clearly. Likewise, volume and duration of speech may signal that the speaker is explaining his ideas in detail, and is presenting his point of view about the subject matter. Such interpretations are grounded in sociolinguistic work related to the way in which speech style specifically (Coupland, 2009; Eckert & Rickford, 2001; Jaffe, 2009) and language style more generally (Fina, Schiffrin, & Bamberg, 2006) reflect both intentional and subconscious aspects of the way in which a speaker positions him or herself within an interaction at multiple levels. These recent accounts build on decades of work beginning with Labov's work on speech characteristics that signal social stratification (Labov, 1966) and Giles' work developing Social Accommodation Theory (Giles, 1984), which describes how speech characteristics shift within an interaction, and how these shifts are interpreted socially. A simplistic interpretation of this work would lead us to believe that hidden within the speech signal are features that enable prediction of social meaning. The Ranganath work (2009) cited above related to detection of flirting supports this view. It is possible to argue that while the essence of transactivity is related to content level distinctions, that it also has a social interpretation, and therefore might be detectable from speech as well. For example, consider that externalizations position students as intellectual leaders within a conversation. However, if true leadership requires that the leader is received as such by the other group members, and transactive contributions indicate that reception, then the occurrence of transactive/ ICC contributions say something about the relationship between speakers. We can then expect that stylistic features that predict positive reception between conversational participants may also predict transactivity/ ICC. The simplest approach to begin such a line of research begins with the types of features used in prior work detecting social aspects of conversations from speech, such as flirting.

Operationalization of the Knowledge Integration Process

When students are working on a given task or a project in a team, they receive a certain amount of information that would help them solve the problem, in the form of a task statement and training materials. In order to solve the given problem, students discuss the materials that are given to them and try to apply them to a potential solution. We are interested in capturing instances when students display reasoning during group discussions that goes beyond what is given and displays some understanding of a causal mechanisms behind the information. Typically some causal mechanism would be referenced in a discussion of how something works or why something is the way it is. In segmenting student talk and identifying which segments display reasoning we are able to quantify amount of reasoning displayed. However, it is important to note that since what we are coding is *attempts* at displayed reasoning, we need to allow for displays of incorrect, incomplete, and incoherent reasoning to count as reasoning. That will necessarily be quite subjective – especially in the case of incoherent explanations. We begin by operationalizing the distinction between non-reasoning statements and reasoning statements, and then we focus on the distinction between reasoning statements that represent new directions within a conversation (i.e., externalizations) from those that build on prior contributions (i.e., ICC).

One important goal in detecting the knowledge integration process is to distinguish instances when students are making their own reasoning explicit from ones that just parrot what they have heard. In our formulation, we consider the task and training materials provided during the experiment to be "given", and we look for contributions where students go beyond that.

Operationalization Step 1: Reasoning Process

Our formulation of what counts as a reasoning display comes from the Weinberger and Fischer's (2006) notion of what counts as an "epistemic unit", where what they look for is a connection between some detail from the given task (which in their case is the object of the case study analyses their students are producing in their studies) with a theoretical concept (which in their work comes from the attribution theory framework, which the students are applying to the case studies). When they have seen enough text that they can see in it mention of a case study detail, a theoretical concept, and a connection between the two, they place a segment boundary. Occasionally, a detail from a case study is described, but not in connection with a theoretical concept. Or, a theoretical concept may be mentioned, but not tied to a case study detail. In these cases, the units of text are considered degenerate, not quite counting as an epistemic unit.

We have adapted the notion of an epistemic unit from Weinberger and Fischer (2006) because the topic of our conversations is very different in nature. The conversations that we analyzed come from a design exercise where 3 participants are asked to design and build an egg holder together. The egg holder will contain an egg, and should protect it from breaking when dropped from a two story high stairwell. As in Weinberger and

Fisher's (2006) notion of "epistemic unit", we look for a connection between two or more concepts. Unlike in Weinberger and Fisher's operationalization of reasoning, where one of the concepts contains at least one detail from the task and the other is a theoretical concept, in our operationalization both concepts can be of either type. We describe our operationalization in detail below. First, examine a segment of a conversation where we have highlighted the instances of displayed reasoning using italics.

> s1: *i think we'll need only one rubber band because the rubber band is circular. We can just break it off right*
> s3: oh yeah. that's a good idea.
> s2: See what are the weights
> s1: *it is some significant difference*
> s2: *Yeah this is heavier. So this could be on top*
> s3: *yeah cause if we did that then that would fall on the bottom, right? It might spin.*

The simple way of thinking about what constitutes a reasoning display is that it has to communicate an expression of some causal mechanism. Often that will come in the form of an explanation, such as X because Y. However, it can be more subtle than that, for example "Increasing the tension makes the spring springier." The basic premise was that a reasoning statement should reflect the process of drawing an inference or conclusion through the use of reason. Note that in the example with the spring, although there is no "because" clause, one could rephrase this in the following way, which does contain a "because" clause: "The spring will be springier because we will increase the tension."

Concepts

The basic building block of a reasoning statement is a concept. We identified 5 types of concepts relevant for our domain, namely theoretical concepts, prior knowledge, physical system properties, emergent system properties, and goals. For each concept, the definition and an example are illustrated in table 1. The examples in the table are from our dataset described in section 3.1, where students are discussing a best approach to build an egg holder. Note that the "system" in this case is the egg holder, plus any materials that are available for use.

Table 1: Definition and examples for the 5 concepts.

Type	Definition	Example
Theoretical concept	principles (i.e. physics principle) and theories	when an object is falling, the force of impact when it hits the ground can be decreased by slowing down the speed.
Prior Knowledge	information based on common sense	Using a small amount of tape would not be enough to hold two bowls together
Physical system properties	elements and characteristics of elements that are available for the system	paper bowl is round, straws are flexible
Emergent system properties	characteristics of elements that appear in a process	stability of an egg holder which emerges as a result of using certain materials
Goal	general believes/ perspectives, anything associated with strong expectations related to points of view	aesthetics of an egg holder, i.e. trying to make the egg holder aesthetically pleasing

Relationship

The presence of multiple concepts in a statement by itself does not determine whether a statement contains reasoning. Rather, the relationship between multiple concepts is the determining factor. For example, a simple list of concepts (e.g., this cup is round, and it is also white) is information sharing, and not reasoning. We identified two types of relationships that signal a reasoning process; 1. Compare & contrast, 2. Cause & effect.

1. Compare and contrast, tradeoff: When the speaker compares two concepts, the speaker is making a judgment, which involves thinking about how two concepts are related to another.
   - The speaker compares two materials ("that" & "rubber band") for his solution: *"I am thinking that might work better than a lot of rubber bands."*

2. Cause and effect: When the speaker uses a cause-and-effect relationship, this process involves establishing the relationship between two concepts through a reasoning process.

The general relation in this category is "doing x helps you achieve y" There are three main types of causal relationship a)cause and effect b)in order to c)analogy. Examples for each of the three types are illustrated below.

- Let's do A because of B: *"Let's use bubble wrap because it cushions the fall"*
- Let's do A in order to achieve B: *"Let's use rubber bands for tying the bag onto the bowl."*
- When a speaker makes an analogy, he is making a link due to the similarity between two concepts. Some of the keywords that signal analogies are "like", "as": *"Oh, you're trying to use the bowl as a parachute."*

Operationalization Step 2: Idea Co-construction (ICC) vs. Externalization

Statements that display reasoning can be either Externalizations, which represent a new direction in the conversation, not building on prior contributions, or ICC contributions, which operate on or build on prior contributions. In our distinction between Externalizations and ICC contributions, we have attempted to take an intuitive approach by determining whether a contribution refers linguistically in some way to a prior statement, such as through the use of a pronoun or deictic expression.

Take the sample conversation we used earlier to illustrate the reasoning contribution. The lines marked with an (E) at the end is a contribution that is categorized as externalization, the ones with a (T) are transactive contributions. The first statement by s1 is an externalization since s1 starts a new topic, thus this contribution is not building on a prior contribution. Subsequent reasoning contributions in this discussion are coded as transactive because they each build on statements that directly precede them.

s1: *i think we'll need only one rubber band because the rubber band is circular. We can just break it off right (E)*
s3: oh yeah. that's a good idea.
s2: See what are the weights
s1: *it is some significant difference (T)*
s2: *Yeah this is heavier. So this could be on top (T)*
s3: *yeah cause if we did that then that would fall on the bottom, right? It might do some spinning. (T)*

Reliability of Annotation

Two coders were initially trained using a manual that describes the above operationalization of reasoning displays and ICC in detail along with an extensive set of examples. After each coding session, the coders discussed disagreements and refined the manual as needed. Most of the disagreements were due to the interpretation of what the students meant rather than the definition of reasoning itself. Therefore, later efforts focused more on defining how much context of a statement could be brought to bear on the interpretation and how. In a final evaluation of reliability for reasoning process, we calculated kappa agreement of 0.67 between two coders over all the data. After calculation of the kappa, disagreements were settled by discussion between the two coders. The coding manual for detecting instances of ICC and externalization is still under development. Our initial round of coding yielded a kappa value of 0.64. The data used in this paper was coded by one coder who has experience with coding for externalization and ICC using a corpus from a different domain.

Automatic Assessment of Reasoning Processes

The purpose of our investigations with speech technology that we report in this paper was to determine the extent to which it is possible to use current machine learning technology paired with simple signal processing preprocessing techniques to distinguish between nonreasoning statements, externalizations, and ICC contributions. We first describe our approach. In the subsequent section, we detail our promising results.

Methods

Our technical approach consists of four main stages: collecting audio data, preparing audio data by transcribing and segmenting the recordings, extracting features from segmented recordings, and displaying predicted scores on a report by applying machine learning. The overall process is detailed in the following subsections.

Collecting the Audio Data

Our corpus was collected in a laboratory setting while students worked face-to-face in groups of three. In this paper, we focus on a subset of the data that has already been collected, transcribed and annotated. The specific task the students are engaged in is to design a contraption to protect an egg when falling the distance of two flights of stairs. This task involves applying a variety of principles of physics. The data we focus on is a 30 minute discussion portion of each 3-student group work session when the participants were designing and building the egg holder together. In order to collect clean speech with each student on a separate channel, each student wore a directional microphone. Nevertheless, although it is possible to clearly identify the main speaker from an audio file, crosstalk, which is the other participants' voices, could still be heard in the background.

Transcribing and Segmenting the Audio Data

For each audio file, the main thirty-minute discussion sessions were transcribed and manually segmented for further analysis. A total of 8 meetings were collected, transcribed, and segmented according to the following two rules. The resulting data contained a total of 4361 segments.

1. Begin a segment when the main speaker starts talking. If there is silence at the beginning of the file when the main speaker is silent, this means that there will be an "empty" segment in the beginning.
2. A segment should contain the main speaker's continuous speech. If there is an interruption (silence or crosstalk) that lasts for more than 1 second, a new segment should be created. When you create a new segment, there should be two boundaries – one that marks the end of the main speaker's first utterance, and another that marks the start of the next utterance after the pause.

Extract Features from Segmented Recording

After the stage of segmenting the data into units, the next stage involved transforming each segmented unit into a set of feature-value pairs. For the feature set, three types of features were extracted. 1. Acoustic features, 2. Phoneme features, and 3. Auxilary features. All three feature sets reflect "how" the words are spoken rather than the content of the words.

Acoustic features capture certain structural aspects of speech such as amplitude, pitch and energy. More intuitively, these features reflect the intensity and energy level of a given speech segment. For instance, a higher value of amplitude means higher volume of the speaker. If there is variation in the amplitude, this indicates that the speaker's volume varied over time. We collected 4 amplitude features, which are the mean value of amplitude over the whole segment, as well as the mean, median, and variance of the 1 second windows in a given speech segment. Similarly, we extracted 4 pitch and 4 energy features: pitch/ energy of the overall segment, mean, median, and standard deviation of pitch over 1 second windows in a given segment. The pitch features were extracted using the YIN algorithm (De Cheveigné & Kawahara, 2002). In addition to these 12 features, 28 of 40 Mel Frequency Cepstral Coefficients (mfcc) were used in the feature set. The initial 40 mfcc features are the result of applying a set of 40 standard filters, which are available as part of VoiceBox Matlab Toolbox (Voicebox, 2010). The mfccs are standard acoustic features that are commonly used in speech processing. They reflect the distribution of energy level in the given speech. Because using all these 40 features would capture somewhat redundant information, we took the top 28 features using principal component analysis (PCA). The decision to take 28 features was based on a rule of thumb that this number of features is sufficient for a variety of speech classification tasks of a roughly similar nature.

Phoneme related features are based on English phonemes, which are the smallest building block of sound in English that carries linguistic meaning. For instance, the phoneme that distinguishes the words tip and dip are the [t] and the [d] phonemes. Sphinx (CMUSphinx, 2010), a speech recognition system developed at Carnegie Mellon University, identifies 48 phonemes in the English language. Thus, we used the 48 phoneme probabilities as part of our feature set. We believe that using phonemes could capture certain aspects of content that would reflect the coding process used by human annotators or the structure of the language data. For instance, according to our operationalization of a reasoning statement, cause and effect relationships can be used to causally connect two concepts. Certain words, such as "because" or "for" are often used in cause and effect relationships. Therefore, phonemes such as [b] or [f] may occur frequently in statements that contain reasoning contributions. In addition to the phonemes, a phoneme-count feature and phoneme rate were computed. The phoneme-count feature shows the total number of phonemes, which tells us how much the speaker spoke in the given segment. The phoneme rate feature is the number of phonemes divided by length of the segment, and provides us with an estimate of how fast a person spoke.

In addition to the acoustic and phoneme features, three additional features were computed, namely duration of the segment and a speaker feature, and a feature that reflects stylistic language matching. The duration of the segment was the length of the given segment in seconds. The speaker feature was a binary feature, 0 if the speaker of the given segment is same as the speaker of the last segment, 1 otherwise. For the feature that reflects the stylistic language matching, we computed the Kullback-Leibler distance between phoneme probabilities, which is a measure of how different two distributions are from one another.

Once all the features are extracted, we used the Adaptive Boosting machine learning algorithm (Freund & Schapire, 1995) to train a predictive model and then evaluate whether it was possible to automatically assign segments of speech as containing a "non-reasoning/ externalization/ ICC" contribution with high enough accuracy. The Adaptive Boosting algorithm was designed to be resilient to noisy data and outliers because of the way it trains a model over multiple iterations, and the instances that are misclassified in early iterations receive more attention in the subsequent rounds through a reweighting mechanism.

Displaying Prediction on Report

The ultimate goal of our work is to use the automatic predictions to provide reports to instructors about how collaborative groups are interacting with one another. Thus, once we obtain the frequency of non-reasoning/ externalization/ ICC contributions from a given meeting, we can display the numbers in a graph so that the instructor can get a sense of which groups need support.

Results

Recall that prior to applying machine learning, human annotators manually labeled the data and verified the reliability of the coding process. Next, we used machine learning to produce labels for the data. Table 2 shows how accurate the machine produced labels are compared to the human labels. We achieve an F-score of .56 for distinguishing reasoning from non-reasoning statements. Distinguishing ICC and Externalizations from other statements is lower, at .35 and .32 respectively. Although the recall and precision rates may not seem very high, they are a significant improvement over a baseline (majority class). For all three types of prediction, duration of segment was the top indicator for determining whether a contribution contained reasoning/ ICC or not. In all cases, the length feature was the most important. This result matches the heuristic that if a contribution contains reasoning, it is longer because the speaker needs time to express his thoughts.

When applying machine learning, we took careful steps to avoid the evaluation results being inflated due to overlap in speakers between train and test sets. Namley, we separated the data into two sets, each with a distinct set of students; specifically, a training set for building a model and a test set for testing the accuracy of the model. Given that we had a limited amount of data, we adopted a 10 fold cross validation methodology where we average the performance obtained for each of the ten test sets. For each test set, 1/10 of the data is set apart as test data, and the remaining 9/10 of the data is used to build a model.

Table 2: Machine learning experiment results showing baseline, recall, precision, F-score, and top 3 most predictive features for the prediction of reasoning, ICC, and externalization statements.

Prediction	Baseline F-score	Recall (%)	Precision (%)	F-score (%)	Feature #1	Feature #2	Feature #3
Reasoning vs. other	0.20	0.63	0.51	0.56	Length (27.8%)	Phoneme rate (6.5%)	12th PCA feature (5.2%)
ICC vs. other	0.12	0.72	0.24	0.35	Length (17.8%)	Phoneme 'B' (18.2%)	12th PCA feature (6.7%)
Externalization vs. other	0.08	0.70	0.22	0.32	Length (35.2%)	2nd PCA feature (4.7%)	9th PCA feature (4.3%)

Conclusions and Current Directions

In this paper, we presented our work towards automatic detection of reasoning displays and ICC contributions in speech data. The need for a tool that presents level of ICC contributions has been demonstrated in our previous study where we investigated the needs of instructors who teach project courses (Gweon et al, 2011). The goal of this paper was to develop technology to address such needs. To this end, we have begun with a simple technique, adapted from other stylistic speech classification tasks. Our work shows promise in that 1) humans can distinguish reasoning and non-reasoning statements with acceptable reliability, although reliability on distinguishing externalizations from ICC contributions needs more improvement, and 2) using machine learning, classification of a statement as reasoning/ non-reasoning is feasible, even with limited training data. Results at distinguishing ICC contributions from others are still weak, especially with respect to precision.

In our future work, more sophisticated adaptations of sociolinguistic work might suggest follow-up techniques. Other pieces of work on sociolinguistics of speech style emphasize social interpretations of stylistic shifts within an interaction (Eckert & Rickford, 2001). For example, Social Accommodation Theory (Giles, 1984) emphasizes the important function of stylistic convergences between speakers within an interaction. This work suggests that more complex features computed over patterns of the types of acoustic and prosodic features that we begin with in this paper may be more conducive to high levels of accuracy. In addition to investigating frequency counts and patterns, we also plan to investigate sequencing and timing rather than just quantity as adopted by Kapur and colleagues (2009). In terms of data, we are currently collecting and annotating audio data from additional meetings as well as other contexts to validate our result further as well as testing its generality across a wider variety of student groups. In addition, because the automatic predictions are not perfect, we must also explore how to properly signal instructors about the confidence level of the predictions.

References

Ai, H., Kumar, R., Nguyen, D., Nagasunder, A., & Rosé, C. P. (2010). Exploring the Effectiveness of Social Capabilities and Goal Alignment in Computer Supported Collaborative Learning, *Proc. ITS.* 134-143.

Ang, et. al. (2002). Prosody based automatic detection of annoyance and frustration. In Proc. *International conference spoken language processing,* Denver, Colorado, USA, 2002, pp. 2037– 2039.

Berkowitz, M., & Gibbs, J. (1983). Measuring the developmental features of moral discussion. Merrill-Palmer Quarterly, 29, 399-410.

Chen, M. (2003). Visualizing the pulse of a classroom. In Proc. MM', ACM Press, 555-561.

CMUSphinx Wiki. (2010). http://cmusphinx.sourceforge.net/wiki/

Coupland, N. (2009). Style: Language Variation and Identity: Key Topics in Sociolinguistics, Cambridge University Press.

De Cheveigné, A., & Kawahara, H. (2002). "YIN, a fundamental frequency estimator for speech and music," The Journal of Acoustical Society of America, 111(4), 1917-1930.

De Lisi, R., & Golbeck, S. (1999). Implications of Piagetian Theory for Peer Learning, A. O'Donnell & Alison King (Eds.) *Cognitive Perspectives on Peer Learning*, Lawrence Erlbaum Associates, New Jersey.

DiMicco, J., et. al. . (2004). Influencing group participation with a shared display. *Proc. CSCW*, 614-623.

Eckert, P., & Rickford. J. (2001). Style and Sociolingusitic Variation, Cambridge University Press.

Faidley, J., et. al. (2000). How are we doing? Methods of assessing group processing in a problem-based learning context. In Evensen, D. H., and Hmelo, C. E. (eds.), *Problem-Based Learning: A Research Perspective on Learning Interactions*, Erlbaum, NJ, 109-135.

Freund, Y., & Schapire, R. E. (1995). A decision-theoretic generalization of on-line learning and an application to boosting. European Conference on Computational Learning Theory. 23-37.

Fina, A., Schiffrin, D., & Bamberg, M. (2006). *Discourse and Identity*, Cambridge University Press

Forbes-Riley, K., & Litman, D. (2009). Adapting to Student Uncertainty Improves Tutoring Dialogues. *In Proc AIED* 2009: 33-40.

Giles, H. (1984). The Dynamics of Speech Accommodation, Amsterdam: Mouton.

Gweon, G., Kumar, R. Rosé, C. (2009). GRASP: The Group Assessment Platform, *In Proc CSCL.*

Gweon, G., et. al. (2011). A Framework for Assessment of Student Project Groups Online and Offline. Puntambekar & Hmelo-Silver (Eds.) Analyzing Interactions in CSCL, Springer. 293-317.

Jaffe, A. (2009). Stance: Sociolinguistic Perspectives, Oxford University Press.

Joshi, M., & Rosé, C. P. (2007). Using Transactivity in Conversation Summarization in Educational Dialog. Proceedings of the SLaTE Workshop on Speech and Language Technology in Education

Kapur, M., & Kinzer, C. K., (2009). Productive failure in CSCL groups. *In Proc, CSCL 2009*. 4: 21-46.

Kay, J.,et al. (2006). *Wattle Tree: What'll It Tell Us?*, University of Sydney Technical Report 582, January 06'.

Kim, T. Chang, et. al. (2008). Meeting Mediator: Enhancing Group Collaboration with Sociometric Feedback, *In Proc. of CSCW,* San Diego, CA, 457-466.

Kumar, R., Rosé, C. P., & Litman, D. (2006). Identification of Confusion and Surprise in Spoken Dialogusing Prosodic Features , *Proceedings of Interspeech 2006*.

Labov, W. (1966). The social stratification of English in New York City, Washington DC: Center for Applied Linguistics.

Liscombe, J. Venditti, J, & Hirschberg, J. (2005). Detecting certainness in spoken tutorial dialogues. *In Proc. Interspeech 2005*. 1837-1840.

Pianesi, F., et al. (2008). Multimodal support to group dynamics. *Personal and Ubiquitous Computing.* 12(3), 181-195.

Pon-Barry., et al. (2006). Responding to Student Uncertainty in Spoken Tutorial Dialogue Systems. IJAIED 16(2)

Ranganath, R., Jurafsky, D., & McFarland, D. (2009). It's Not You, it's Me: Detecting Flirting and its Misperception in Speed-Dates. *Proceedings of EMNLP 2009.* 334-342.

Rosé, C. P., et al. (2008). Analyzing Collaborative Learning Processes Automatically: Exploiting the Advances of Computational Linguistics in Computer-Supported Collaborative Learning , IJCSCL. 237-271

Soller, A., & Lesgold, A. (2003). A computational approach to analyzing online knowledge sharing interaction.In Proc. AIED, 253-260.

Teasley, S. D. (1997). Talking about reasoning: How important is the peer in peer collaboration? In L. B. Resnick (Eds.), Discourse, tools and reasoning: Essays on situated cognition, 361-384.

Voicebox (2010). http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html

Weinberger A., & Fischer F. (2006). A framework to analyze argumentative knowledge construction in computer supported collaborative learning. Computers & Education; 46, 71 – 95.

Acknowledgments

This research was supported in part by NSF EEC grant number 064848.

Learning through Collaborative Creation of Knowledge Objects in Teacher Education

Crina Damşa, Sten Ludvigsen, InterMedia, University of Oslo, Postboks 1161 Blindern 0318 Oslo, Norway
Email: crina.damsa@intermedia.uio.no, s.r.ludvigsen@intermedia.uio.no

Abstract: This contribution presents an empirical study of object-oriented collaboration. The participants are groups of teacher students, who work on knowledge objects (e.g., didactic materials, study materials and guidelines for teachers) intended to address problems identified at schools where they pursue their internships. The knowledge creation approach to learning, which places collaborative creation of knowledge objects at its core, served as a guide for the pedagogical design, which includes specifications of the collaborative activities and tools utilized. We collected different types of data. Analyses followed three lines of investigation: analysis of group interactions, of concepts and ideas uptake, and of object co-construction and development. Findings show various degrees of idea sharing and of co-elaboration of object iterations. These findings assist us in formulating recommendations for future research and pedagogical design, especially with regard to pedagogical settings, and technological support for object co-construction and co-elaboration.

Introduction

Exposing students to knowledge practices and the problems they will face as professionals is a challenging task for higher education institutions. Various studies (see Muukkonen & Lakkala, 2009) posit that collaborative settings and problems with open-ended character entice students to activate and employ knowledge acquired in different settings, and to envision solutions to these problems together with peers. This calls for theoretical and practical knowledge to be concretized in knowledge objects wherein this knowledge becomes materialized (e.g., in designs, research reports, educational material, etc.) and transparent for the participants involved. However, becoming actively involved in such complex processes and creating sophisticated knowledge objects is a challenging task for students. In order to learn more about how students can be supported in their endeavor, in-depth examination of these aspects is needed. In this paper we investigate how teacher students collaborate to create and develop knowledge objects that will be employed at the schools where they pursue their internships. In our analyses we focus on the interactions that constitute the object-oriented collaborative process, and on the mechanisms of iterative knowledge object development.

Various theoretical perspectives (Bereiter, 2003; Vygotsky, 1987) reconceptualize learning as a social activity, and knowledge as socially constructed. The idea of learning as a social activity has been explored and investigated from various angles, i.e., the situated approach or the different versions of social-cultural theory. In our work we employ the latter, and we emphasize the epistemic and regulative dimensions of learning in interaction and collaboration, aimed at creating new knowledge objects. The focus on creation and development of knowledge objects is proposed by Paavola and Hakkarainen (2005), as part of their view of learning as a collaborative, knowledge-creating activity. As posed by Van Aalst (2009), this view goes beyond objectivist stances, and emphasizes that understanding and knowing are mediated by the objects that are created and shared by a community. Rather than residing inside individual minds, ideas are regarded as cultural objects that mediate learning and understanding. In general, objects play an important role in human interactions, as they structure interaction, generate problems (Jordan & Henderson, 1995), or provide groups with motives for interaction or outcomes to reach for (Stahl, 2006). However, in spite of the theoretical account of objects in learning settings, not much is known about how they emerge from these interactions, about the way these objects develop and change over time, or about how they are mediated by collaborative technologies. When studying objects, the emphasis has been placed mostly on the role of objects (see Eckert & Boujut, 2003) rather than on the nature of the objects themselves (Ewenstein & White, 2009). Hence, in order to investigate whether objects play a key role in knowledge creation and learning, we need to understand how objects become constituted and how they are developed.

This paper attempts to examine the mechanisms of collaborative creation of knowledge objects in a teacher training program. The aim of this study is twofold. First, we aim to pin down interactions that lead to knowledge object development. Second, we seek to understand how the knowledge objects created gain shape through collaboration, and how object development can be related to interactions between participants. In the case study reported in this paper, we employ a qualitative perspective, as well as methods to gain a better understanding of the aforementioned processes and to depict their characteristics. A set of data from an extensive design study that involved teacher students working on collaborative projects is used to illustrate these mechanisms. Analyses aim to unveil the characteristics of learners' interactions that contribute to object advancement.

Productive Interactions and Object Development

The idea that interactions are productive builds on the concepts of social and constructive interaction. The former accounts for learning as a socially shared and culturally situated phenomenon wherein activities and meanings are subjectively realized (see Ludvigsen, 2009). The latter refers to learning in and through dialogue (Baker, 2009), and is related to the social construction of meaning and knowledge as proposed by the various socio-constructivist views of learning. In knowledge creation, the denotation of this concept expands from collaborative meaning-making to the materialization of knowledge objects.

A brief literature study shows that there are no clear-cut definitions of the concept of "object". Objects are referred to as a collection of artifacts that individuals work with, i.e., create, measure, and manipulate. To avoid confusion between the terms "object" and "artifact," we refer here solely to objects, given the synonymous use of both terms in the literature (see Paavola & Hakkarainen, 2005). It seems plausible to adopt a broad definition of objects, most notably because they also vary in their degree of abstraction (Ewenstein & Whyte, 2009). We adopt the distinction between historical object of activity, which is developed and represents an answer to societal needs (e.g., assessment systems), and situational object, which is a manifestation of the historical object but is procedural and discursively constructed (e.g., a report produced by a group of students in a project). The interaction of learners, in this case, is only understandable in relation to this situational object they construct together (Jahreie, 2010). Thus we employ the concept of object as conceived in the latter stance.

Knowledge objects, also referred to as epistemic objects or epistemic things (Miettinen & Virkkunen, 2005), are considered to focus on what is not yet known. These authors described epistemic objects as open-ended projections oriented towards something that does not yet exist, or to something that is not known for sure. These objects are characterized by their incompleteness and continuous evolution (Ewenstein & Whyte, 2009), and as projects under construction, moving from potential raw material to meaningful forms (i.e., iterations), resulting in a final outcome. Work itself is therefore a continuous process of transforming an object from its current state into a required end state. Examples of epistemic objects may include theories, plans, protocols, etc. Our stance becomes instrumental when attempting to understand the way objects are created and how they develop. We approach these processes with the premise that knowledge objects are evolving entities, which are shaped by the interactions around them, but which also shape those interactions. This stance allows us to understand how students' learning evolves over extensive periods of time.

Methods

This one-year study was conducted at a university of applied sciences and teacher education in the Netherlands that prepares pre- and in-service teachers for secondary vocational education. The curriculum is based on courses called Professional Situations (PS). At the start of this project, learning in the PSs was organized around a combination of lectures and seminars given by the teachers, and individual student projects. These projects required students to perform various learning tasks (literature reviews, risk assessments, designs). Twenty groups of 73 mixed-age students, enrolled in three randomly selected PSs, participated in our study.

Through the project we present here, students' learning was enhanced by facilitating participation in collaborative projects in the PSs. Groups of students were required to develop and report on knowledge objects (e.g., didactic materials, study material for teachers, guidelines for applying new teaching methods, etc.). Various scaffolds were provided, such as workshops on object-oriented collaboration, templates for work plans, and tool-training sessions. Brainstorming sessions were organized to discuss topics collected at those places where students pursued their internships. Next, topics were clustered and, based on these clusters, project topics were formulated by the students themselves. Once the project groups were formed, students' interactions focused on the chosen topic, and took place mostly within group and with the tutor. Group members had the opportunity to work online, to meet face-to-face, and to have regular coaching sessions with the tutor. Groups were required to provide theoretical grounding for their products and, at the end of the project, students were to reflect on the quality of the object and their individual and group activities.

Technological support was provided by the Knowledge Practices Environment (KPE, see www.knowledgepractices.info), a web 2.0 application that provides virtual collaboration spaces. In the KPE, each group had its own shared virtual work space. Inside these spaces, students were expected to use functionalities that supported the organization of the collaborative process (i.e., task creation and planning functionalities) and the creation and iterative development of knowledge objects (i.e., item creation, versions, commenting, linking, sources collection and display, and chatting). We collected a rich set of data, consisting of the following categories: (a) interaction data (recordings of group meetings, e-mail correspondence, field notes); (b) knowledge objects, both produced in KPE (documents, wiki, notes, comments) and during fieldwork (meeting notes, comments, course documents, project material), and (c) reflection data (reflective questionnaires, semi-structured interviews). This contribution discusses data drawn primarily upon the transcribed group discussions, object iterations (text documents), and meeting and researcher's notes. Data were chronologically ordered, and the analyzed recordings were transcribed verbatim. We used a set of analytic tools embedded in the KPE, which provide an overview of the members' activities in the shared space and contributions to the developing objects.

For the purpose of this study, a *case* was defined as the activities and products of one group of students during the 20-week course period. The analyses followed three lines of investigation. *First*, we selected and coded action-relevant episodes from the group's interactions, in order to understand the nature of their activities. We conducted a focused examination of those episodes and selected from the five discussion sessions, equally distributed over the project period. The categories employed were based on a coding system developed by Damşa, Kirschner, Andriessen, Erkens, and Sins (2010), and refined based on the models of discourse by Van Aalst (2009). The case emerged based on a literature review and iterative analyses of other (similar) data sets. The resulting categories distinguish between epistemic actions aimed at eventually creating knowledge (i.e., creating awareness, sharing knowledge, creating shared understanding, and generative collaborative actions) and regulative actions (including planning, coordinating, monitoring, and reflecting activities). We labeled all actions and traced the newly introduced concepts and ideas. *Second*, we mapped the relevant concepts and ideas discussed during the analyzed interactions, and examined whether they were adopted and materialized into the knowledge objects (and their iterations). The relevant-concepts map was created based on the first discussion session of the group, then it was subsequently revised. Only new relevant-concepts and ideas were added to the map, based on subsequent group discussions. Next, we created a concept map based on different versions of the knowledge object the group created. We compared the concept maps created based on interaction and object versions and established the similarities, the relevant-concepts and ideas that were put forward most frequently and those that were elaborated in the final object version. *Finally*, we operationalized object progression and elaboration, and analyzed how the relevant concepts and the ideas generated were gradually materialized and elaborated into object iterations. Several readings of these iterations led to the understanding that object development can be described based on two aspects: first, the way concepts and ideas are adopted from one interaction to another and, second, how ideas and concepts are conceptually expanded from one iteration to another, i.e., elaboration. To determine the degree of elaboration, we employed an instrument containing four levels, derived from Cummings, Schlosser and Arrow's (1996) concept of the integrative complexity of ideas. The following levels were distinguished: (0) idea or concept was not further elaborated in the subsequent object iterations; (1) low integrative complexity means that the content is constructed using (literally) information selected from one or more sources; (2) moderate integrative complexity means that multiple perspectives are expressed, based on information sources and one's own insights, but without new conceptualizations; (3) high integrative complexity specifies a dynamic relation between perspectives, including integration of source-based knowledge and information, and one's own ideas and interpretations, which can lead to new conceptualizations.

Empirical Analyses

The analyses concern one student group, named BOKITO, consisting of four female students; however, the whole data corpus was analyzed, following an ethnographic approach, to identify variations between groups and aspects that reflect problems illustrative of students' experiences. We selected this group because their activities and knowledge object provide a good illustration of how knowledge is created. This sampling can be situated between representative and theoretical sampling. We analyzed the group's materials and activities over the course of 20 weeks (February–June 2010). This group's members were enrolled in PS11, which focuses on the analysis, design, and evaluation of assessment instruments. Students were required to conduct a literature study and, based on its conclusions, design an assessment-related product. The group conducted a study on quality criteria for authentic assessment methods in competence-based education, and developed a set of guidelines for teachers who intend to use such methods in secondary and tertiary vocational education. The product was concretized in a report containing the literature study, a description of the developed object, and explanations regarding the choices they made in the design process. Additionally, the group summarized these recommendations in artifact resembling a fan, with several plasticized sheets, containing the main guidelines.

This section begins with a discussion of an interaction sequence, followed by an illustration of the uptake of relevant concepts and ideas put forward in the respective interactions. Next, illustrations of the object progression and elaboration are presented and discussed.

Productive Interactions

In this group's work we noticed a high concern among students about understanding concepts. Concepts related to *assessment* took a central place in group discussions. A relevant portion of the interactive sessions was dedicated to deciding which concepts and ideas were relevant, what their meaning was, and how they should be explained, linked to each other, and elaborated. One strategy for elaborating such conceptual systems was by discussion; another was by searching for external information and integrating it with their own ideas.

An analysis of sources led to the development of iterations of theoretical section A, in which the group analyzed and separately presented both traditional and new quality criteria for assessment, and three assessment methods specific to competence-based education. The following step was to shape the next section of their knowledge object, wherein they planned to establish a relationship between the three assessment methods and the quality criteria. The following discussion episode is selected from an interactive session used for brainstorming and making plans and establishing agreements for this object section.

Excerpt 1, eighth project week: Elaboration of concepts and object versions

Participant	*Discussion line*	*Analytic categories*
1. Liz:	…So, in Part A we have these theoretical sections, wherein the main criteria validity, reliability, transparency, and authenticity are defined and explained…and then …	(3) Structuring concepts
2. Ann:	Then we apply them. No, wait…	
3. Jane:	No, we don't apply them. **We establish the relationship between the criteria and each method.**	(3) Framing focus
4. Ann:	Yes, that's it. In here **it goes from broad to more specific, related to the practice situation and the specificity of the method**. […]	(3) Elaborating ideas
5. Liz :	**But how do you translate the theoretical idea to the very practical way** these assessment methods are used in our schools?	(3) Problema-tizing
6. Ann:	A good point, but we shouldn't go in that direction. I see it differently. I would suggest that **we take each method as described and discuss how we think it should be applied to meet the quality criteria.**	(4) Generating new idea
7. Jane:	But that is again very theoretical, don't you think? Why not **take the standard way designed for our schools and check that out in relation to the criteria? Then make conclusions** based on that.	(4) Negotiating ideas
8. Ann:	Can you explain? I don't get it.	
9. Jane:	Errrm, well, **in my school they have, for example, this common assessment form for proof of competence**. All the assessors must use it, and the assessment of the same pupils by **more assessors** is discussed at the end. It's something that has to do with **reliability**.	(2) Sharing information (3) Framing Focus
10. Liz:	... While reading about validity I came across things I recognized in practice. **The task, for example. How that is covered by the tests has to do with content validity**.	(2) Sharing information
11. Ann:	Also with authenticity. Tasks have to reflect the real work situation.	(3) Elaborate idea
12. Liz:	Right, see? I guess this is the way we can go.	
13. Jane:	And the **context,** too, can influence reliability.	(4) Sharing inf.
14. Ann:	Yes, **let's see how we approach this**. I guess **we can each collect the information from own school**, maybe **from the** web about other schools, too? **Then we discuss and then start writing**.	(5) Planning and coordinating process

This episode shows the way group members attempt to establish a link between the theory on quality criteria and the existing methods. The first utterance by Liz (line 1) immediately provides a good summary of the relevant quality criteria the group chose to elaborate (*validity, reliability, transparency*, and *authenticity*). At this point, the existing knowledge is structured, and it serves as a basis for brainstorming about the conceptual links possible in the respective object section. In the next lines, the students attempt to establish the focus of this section, and they do it by taking up the idea put forward in their second meeting (line 3: *We establish the relationship between the criteria and each method*). By problematizing and elaborating on the proposed idea, students come to a better understanding of what this relationship actually means and how they might approach it. This is needed because, while the general goal of this step (i.e., establishing the relationship) is clear, the way they will elaborate on this relationship still needs to be delineated. By generating a new idea (line 6), by negotiating aspects that seem not to suit some of the members' views of the conceptual system under consideration (line 7), and by elaborating on their own or others' ideas (lines 4, 11), the students reach a common understanding of the conceptual system under development. Based on this, they pin down the strategy they will use in order to materialize the ideas put forward.

Uptake - Relevant Concept Map

The concept map created based on the object iterations defined one week after the above interaction shows that the relevant concepts from this interaction sequence are adopted by the group and processed into the object. In the concept map in Figure 1, we clustered these concepts (using interrupted-line rectangular frames) in order to indicate how the concepts were treated by the group, both in the discussions and in the object iterations. In fact, this is a conceptual system created by the group members themselves during verbal interaction and, later, in the text iterations. The gradual shading indicates the importance assigned by the group to the concepts (the darker shading indicates the most important concepts), while figures attached to each box indicate the interaction analytic framework category - see excerpt above). In the center, a number of main concepts are shown (*quality criteria for authentic assessment, authentic assessment methods, relationship methods*, and *criteria*), which were adopted from the previous discussion sessions and the first object iteration. These concepts cover the main topic of the project. In the upper-left corner are the authentic assessment methods proposed in the second meeting, which were adopted by the group as one of the focus of the knowledge object. These methods are defined and described in the first section of the knowledge object, the research report. The two clusters on the right side of the map display the quality criteria as identified by the group members, in discussions (also with their teacher) in the literature, and in the preliminary object iterations. In this map version, the group indicates that “validity” and “reliability” are traditional psychometric criteria, while the others are specifically related to authentic assessment. The group decided on the former two and on “transparency” and “authenticity” to be further elaborated in their knowledge object, and they provided arguments for their decision. The cluster of concepts in the lower section of the map indicates some concrete elements (instruments, activities) used or performed in the educational practice to ensure quality of assessment. These are used by the group in the object iterations, in order to bridge concepts and to emphasize the relationship between the two conceptual dimensions (methods and criteria) elaborated upon. The last cluster, in the lower-left corner, contains the weight points of the knowledge object and the way the relevant concepts were to be elaborated.

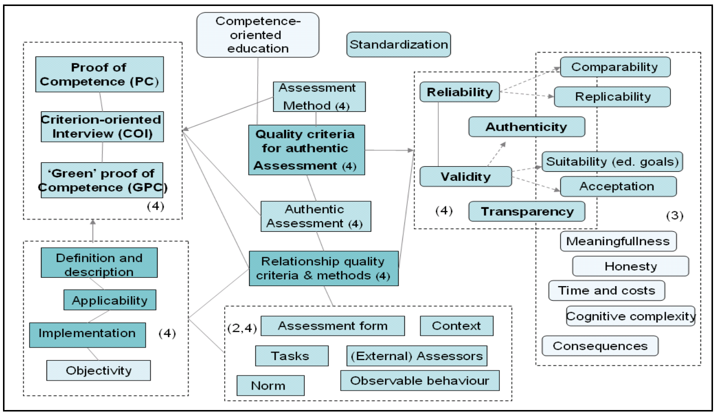


Figure 1. Relevant-concepts Map Based on Interaction in the Eighth Week and Object Iterations in Weeks 9-12.

An analysis of the progression of the knowledge objects shows that the majority of the concepts proposed in the first two group meetings were adopted and evolved into knowledge objects (materialized in the iterations) at later stages. The main concepts (*quality criteria, assessment methods*) were employed throughout the entire project period. These concepts were defined and described in general terms in the introduction of the report, but no extended analysis of these general concepts was conducted. The concepts directly addressed in the project, i.e., the three assessment methods and the four quality criteria, were adopted fully, explained and described based on information sources, and elaborated in relation to each other. Some concepts were dropped, and some interaction instances indicated that this was a deliberate decision of the group, since these concepts were not (directly) linked to the focus of the study.

Object Progression and Elaboration

The general overview based on the analytic tools provides an insight into the individual contributions to the shared knowledge object, in terms of types of items created using functionalities at their disposal in the shared virtual work space. As the figures in Table 1 show, the number of contributions is rather comparable, without obvious outliers. Jane is the only member who seems to have uploaded or modified most files, but the meeting notes and observations indicated that she was the one who was assigned to upload all product iterations when finished. The higher number of links created by Liz indicates that she organized the materials and created the relationship between various sections and iterations of the object.

Table 1: Illustrative overview of contribution to the object.

Group member	Files (iterations)	Links between items	Comments (on files)	Web links
Jane	38	12	7	5
Ann	26	11	5	0
Liz	21	21	6	0
Julie	23	11	8	0
Teacher 1	0	0	8	2

In qualitative terms, the activities emerged according to the general research activities phases. In the beginning, brainstorming took place and ideas were proposed as topics for investigation. Once the focus was decided upon, following discussions with the stakeholders and a preliminary source search, a joint plan was created. The plan consisted of the research topic, focus, and the types of activities to be performed. Once this plan was accepted, the group proceeded with individual source searching and information processing. Information was first summarized, then discussed with the group, then processed individually in text (in object iterations), and later discussed in the group again. Also, group members gave written feedback on the written draftsAll the object iterations were placed in the group's shared space, where there were possibilities for conceptual structuring and commenting. Once the theoretical section on methods and criteria was partially elaborated, work on linking the theory with the practical solutions commenced. The same strategy was applied as in the previous phase. Once various sections were commented upon and finalized, they were organized and integrated into the final object.

Object elaboration at level zero (0) was applicable to 10 concepts identified in the group's interactions, which were then omitted. Other types of data indicate that these concepts were not elaborated, based on deliberate decisions, because they were considered too marginal to the main topic. While the purpose of these analyses was not quantification and normative assessment, we can state that, on average, overall elaboration is situated as the second level of our framework. At this level, information from sources and group's own ideas were used, sometimes combined, in order to explain and expand the concepts in discussion. However, the levels of conceptualization differed between the sections of the report. In section A, the literature study of the assessment methods and criteria, the average level was two (2). This indicates that the literature was not interpreted and processed in depth. In Table 2, we present an example of a comparison between the source text and the text written by the students, which was assigned level 1 of elaboration, a rather low elaboration level. Furthermore, a relatively low number of sources were used when defining and describing the quality criteria. Nevertheless, the group succeeded in structuring the text and employing the source information in a meaningful manner, which contributed to achieving their goal, namely presenting the given concepts in a concise manner.

Table 2: Example of a level 1 elaboration object sequence.

Text from sources	Group's text
2 Comparability Consists of that a Competence Assessment Program (a) must be performed consistently. The (b) assignments done by students, the de-evaluation procedure, and the evaluation criteria must be comparable in sufficient degree for all students. In competence-based education this is (c) not always easy to oversee and it also doesn't means that the entire assessment should be standardized. But a study program can (d) pose particular requirements to the practical training in a company and to (e) the tasks a student must perform as part of this practical training.	Comparability Consists of that the assessment (a) must be performed consistently. The (b) assignments done by students, the de-evaluation procedure, and the evaluation criteria must be comparable for all students. In competence-based education this is (c) not always easy to oversee. It also doesn't mean that for each assessment there is a standard that can be employed, but there are, for example, (d) particular requirements that can be posed to the practical training and to (e) the tasks students must perform during the practical training.

In section B of the product, elaboration of concepts was at the highest level. Information sources were used to explain various theoretical aspects but, in general, the information from the sources was well integrated with students' own ideas. In this section, it was students' interpretations that were the leading element in the elaboration, and not the structures provided by the literature. The former became visible in the way particular ideas were elaborated across the versions, and by the conceptual elaboration techniques that were employed. Such techniques involved: creating arguments based on conflicting input from the literature versus education practice, providing examples, or providing elements of reference for comparison. To illustrate the aspect of gradual elaboration of one's own ideas and materialization into the text, we present an example of how a text section evolved across the three iterations we analyzed. Because the original text was too extended to be presented here, we summarized and synthesized the content (see Table 3). The text deals with the relationship between the method (*Proof of Competence*) and the criterion (*Authenticity*). In each of the three columns, the type of material and the strategy used to work with that material are described, while in the boxes attached to the table we indicate the type of action taken by the group upon the respective iteration. In the first iteration (column 1), the idea of the relationship between the two concepts is in an initial phase of elaboration. Knowledge presented in the previous section is repeated, and information from sources is used to provide arguments for this relationship. Here the students establish an unelaborated link with educational practice. Iteration 2 is further elaborated towards a more integrative use of one's own ideas, information from sources, and information from practice. There are not yet any conceptualizations based on these different perspectives and there is, still, information presented in list form, without explanations. The third iteration, which follows a systematic round of feedback by all group members, shows a higher degree of elaboration and the creation of a new conceptualization that employs different perspectives. Examples are used to support the given explanations.

In general it appears that students adapted their strategy to the purpose of each section. In the first section the goal was to provide definitions and descriptions of the concepts agreed upon in the interaction sessions, not to provide extended elaborations. In the second section, the goal was to establish a link between the quality criteria and the three assessment methods, which required interpretations, conceptual expansion, and application of theory. To conclude, we can state that the level of conceptual elaboration varied rather strongly between sections, which means that the object development was not optimal.

Table 3: Example of elaboration across object versions.

Iteration 1	Iteration 2	Iteration 3
1. Method and criterion – definition and explanation 2. One paragraph containing a summary of source information on why this criterion is important 3. Three sentences on possible assessment situations to be taken into consideration as assessing institution.	1. Method and criterion – definition and explanation 2. One paragraph containing a summary of source information on why this criterion is important 3. Elaborated paragraph explaining several assessment situations to be taken into account by the assessing institution 4. Summary of literature on a framework used to formalize the possibilities and the options the assessing institutions has to meet the criterion 5. List of tools associated with the above framework	1. Method and criterion – definition and explanation 2. One paragraph containing summarizing source information on why this criterion is important 3. Elaborated paragraph explaining several assessment situations to be taken into account by the assessing institution 4. Presentation of the framework using the source information, but processed reformulated) and using quoting 5. List of tools associated with the above framework with explanations for each tool 6. Paragraph on how the tools can be employed and applied in to the given situations presented by point 3; with examples.
Group discussion on the applicability of the criterion	Written feedback on preliminary text by group	Structure of final version

Discussion and Concluding Remarks

This paper focused on understanding the mechanisms of collaborative creation of knowledge objects, by analyzing the interactions of students engaged in authentic projects, and by tracing the development process of the group`s knowledge objects. The analyses and preliminary findings presented in this paper provide an insight into how knowledge objects are created through a collective effort in this type of pedagogical setting. We were able to interpret the interactions as being productive through mapping of concepts and ideas put forward in interaction sequences and by "following" these concepts as they were pursued and as they "grew" through object iterations. The analyzed *interaction*s had a heterogeneous character. The productive interactions that explicitly led to shared agreement and understanding, and to generation of new ideas, enhanced the conceptual development of the shared knowledge objects. The analyses yielded *sets of relevant concepts*. The relevant concepts identified through concept-mapping showed that the majority of concepts and ideas put forward in these productive interactions were adopted, expanded, and elaborated. *Elaborations* led to the materialization of the group's shared views and stance, since the concepts chosen for elaboration were decided upon in the interaction. However, much of the work of expanding these concepts and ideas took place individually and led, at times, to lower levels of elaboration. Finally, individual contributions were closely combined with joint activities. Individual elaboration led to numerous feedback moments by the other group members. The analyses indicated that the object sections that were discussed more than once showed a higher degree of elaboration.

These findings provide ideas for improvements of the pedagogical setting. First, productive interactions should focus on the epistemic aspects of work on the shared object (see also Damşa et al., 2010). This group of students approached their interaction not only from a procedural level, but also interacted around concepts, ideas, and the knowledge object itself. They became involved in desirable practices that should be further stimulated and applied in other settings. Next, we partly illustrated successful instances showing the materialization of the group's views and ideas into concrete objects. Different studies (Van Aalst, 2009) indicate that this is not the default way of working on shared knowledge objects. In most cases, complete division of labor occurs. In this group we also observed deliberate attempts to discuss knowledge domain-related ideas, collaborative editing, and consistent feedback on individually written texts. This indicates an orientation towards a collective way of acting upon the shared object, to be considered when organizing and coaching group work. It is desirable that students work towards high levels of elaboration, but it should not be taken for granted that they are always aware of the differences in the elaboration levels. Hence, these findings call for attention to students' understanding of this pedagogical setting and, also, to how these types of designs can provide more specific scaffolds for students when entering knowledge co-construction processes. At *theoretical* level, our study contributes to the field by, first, problematizing and extending the discussion on the concept of "knowledge object". A second contribution consists of the attempt to link this concept, and the way it is materialized, to the productive interaction concept. This theoretical discussion is taken up at *empirical* level, where we investigate this link by creating and integrated analytic framework that addresses both the interaction process and the developing knowledge objects. Hence, at *empirical-analytic* level, our contribution rests in the attempt to employ different analytic angles and to tackle these processes from an organic perspective. However, while very functional in categorizing interactions, our analytic categories prove to be insufficient in unveiling the way students follow up on the topics based on the interactions. In the same way, the elaboration framework developed based on Cummings et al. (1996) provides us with an ideal model. This model is useful when unpacking the conceptual structure and elaboration of the object but it leaves out important aspects of the collective contributions to and the (gradual) progression of the object. As such, this study contributes to the field by demonstrating the necessity of an integrated analytic approach, and a joint interpretation of the processes involved. Nevertheless, there are several aspects that require careful consideration when attempting to formalize the analysis framework employed. We have used an explorative in-depth case-study approach that attempts to unveil mechanisms of a complex process. In order to refine the method used here and to increase the potential for future generalizations, an extended analysis on a larger sample is recommended. Regarding the design and use of *technology*, this study showed how face-to-face collaboration can be organically combined with online support. Although the analyses in this paper did not focus specifically on the use of technology, the processes and products analyzed emerged in the context of technology-supported collaboration. Reflections on these activities indicated that KPE provided good support for work on developing the knowledge objects (individually or jointly) and organizing the group work, but lesser support for communication.

To conclude, our assumption that productive interactions lead to the emergence and development of knowledge objects worked upon by the participating groups is partially confirmed by the illustrative material analyzed and discussed. Furthermore, the emphasis was on envisioning ways to analyze and display the (interactional) processes taking place and whether there is a follow-up to these interactions in the object iterations and the final knowledge objects. Our attempts to understand and depict the process of collaborative creation of knowledge objects and the way these object develop provided important insight, but more work is needed to specify and ground these phenomena.

References

Baker, M.J. (2009). Argumentative interactions and the social construction of knowledge, in N.M. Mirza & A.-N. Perret-Clermont (Eds.), *Argumentation and Education: Theoretical Foundations and Practices*, Berlin: Springer Verlag, pp. 127-144.

Cummings, A., Schlosser, A., & Arrow, H. (1996). Developing complex group products: Idea combination in computer-mediated and face-to-face groups. *Computer Supported Cooperative Work, 4*, 229-251.

Damşa, C.I., Kirschner, P.A, Andriessen, J.E.B., Erkens, G., Sins, P.H.M. (2010). Shared epistemic agency: An empirical study of an emergent construct*. The Journal of the Learning Sciences*, *19(2*), 143-186.

Eckert, C., & Boujut, J.F. (2003). The role of objects in design co-operation: Communication through physical or virtual objects. *Computer Supported Cooperative Work, 12,* 145-151.

Ewenstein, B., & Whyte, J. (2009). Knowledge practices in design: The role of visual representations as 'epistemic objects'. *Organization Studies, 30*(1), 7-30.

Jahreie, C.F., & Ottesen, E. (2010). *Construction of boundaries in teacher education: Analyzing student teachers' accounts*. Mind, culture and activity ISSN: 1074-9039

Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *Journal of the Learning Sciences, 4*(1), 39-103.

Miettinen, R., & Virkkunen, J. (2005). Epistemic objects, artefacts, and organizational change. *Organization, 12*(3), 437-456.

Muukkonen, H., Lakkala, M. (2009). Exploring metaskills of knowledge-creating inquiry in higher education. *Computer-Supported Collaborative Learning, 4,* 187-211.

Paavola, S., & Hakkarainen, K. (2005). The knowledge creation metaphor – An emergent epistemological approach to learning. *Science & Education, 14,* 535-557.

Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge*. Cambridge, MA.

van Aalst, J. (2009). Distinguishing knowledge-sharing, knowledge-construction, and knowledge-creation discourses. *Computer-Supported Collaborative Learning, 4*, 259-287.

Vygotsky, L. (1978). *Mind in society: the development of higher social processes*. Cambridge, MA: Harvard University Press.

Acknowledgments

This study was conducted in the context of the KP-Lab project (see www.knowledgepractices.info). We thank Patrick Sins and Johan van Strien for ideas regarding the analyses, and our colleagues in the InterMedia department at the University of Oslo for their feedback on the text.

Assessing the Use of a Trace-Based Synchronous Tool for Distant Language Tutoring

Mireille Bétrancourt, FPSE, University of Geneva, CH-1211 Geneva 4, Mireille.Betrancourt@unige.ch
Nicolas Guichon, Université de Lyon 2, Laboratoire ICAR, Université Lyon 2,5, Parvis Descartes, 69342 Lyon cedex 07 FRANCE, Nicolas.guichon@univ-lyon2.fr
Yannick Prié, Université de Lyon, CNRS Université Lyon 1, LIRIS, UMR5205, F-69622, France, yannick.prie@liris.cnrs.fr

Abstract: This article presents a pilot study carried out to investigate the potential of a functionality marker setting, included in a synchronous collaborative videoconferencing platform (VISU). Markers, supported by a trace-based system, are designed to facilitate tutors' activity. They provide tutors with (1) the possibility of annotating their distant learners' learning activity, and (2) information pertaining to their own behavior during pedagogical interaction, which can potentially enhance their professional performance as online language tutors. This study concentrates on the marker-based traces of eight language tutors collected in the course of pedagogical interactions with their distant learners during a seven-week transnational collaborative project. It presents both quantitative and qualitative analyses of the use of markers during synchronous language teaching sessions and assesses the utility and usability of such a functionality for language tutoring in order to inform future design and training.

Introduction

While asynchronous collaborative distant learning is now relatively well researched, synchronous distant learning is less well documented although it is gaining more and more attention thanks to the availability of efficient technologies for synchronous communication (videoconferencing, chat, whiteboards, screen sharing, etc.). This research was conducted in the context of the ITHACA[1] project, an interdisciplinary (language education, computer science, cognitive science) project whose aim is to study how traces can be used for providing such synchronous collaborative systems with reflective features so as to enhance their functionalities in the context of activities involving tutoring at a distance. We define traces as features of the system that offer the means to provide instant or delayed feedback regarding the unfolding of the pedagogical interaction, and thus allow both learners and tutors to reflect on the experience.

This paper is organized in two parts. First, we will present the rationale and the design of a synchronous trace-based prototype tool called VISU2 featuring means to share and interact with three types of traces: automatically recorded interaction traces; marker-based traces composed of *ad hoc* annotations added by the user in the course of the activity; comment-based traces composed of *post hoc* annotations added by the user in order to comment upon his/her past activity. In the second section of the paper, we will focus on the use of one type of trace - marker-based traces – as used during a real 7-week long real distant pedagogical interaction between language tutors in France and distant learners from a North-American university. Our results, which are mostly qualitative, tend to support the potential of such traces for online language teaching as well as language teachers' professional development. The article is organized as follows: We begin with a brief overview of distant tutoring and trace-based reflexivity, before examining the VISU tool in section 3, the experiment in section 4 and the results in section 5.

Theoretical Rationale

Tutoring in Synchronous Distant Learning

Synchronous collaborative learning systems support more or less predefined collaborative activities (Lonchamp, 2006), with various roles for the users and various tasks to be carried out. When synchronous online tutoring is involved, teachers are often faced with an ill-defined pedagogical situation and have to develop new competencies and attitudes that are necessary for effective interactions with distant learners in addition to more traditional pedagogical skills (see Guichon, 2009). Besides, conducting a synchronous online interaction induces a high cognitive cost because it involves carrying out several tasks at the same time as well as managing interrupted connections and various distractions (Oviatt et al., 2004). Social and Cognitive Awareness and Navigation (SCAN) tools (Buder et al., 2009) can provide feedback and a sense of group awareness to participants of the interaction. The rationale that underpins this research is that teaching (and teacher training) can be facilitated with trace-based SCAN systems that foster teachers' reflexivity. Such systems help augment their awareness by providing them with traces that are pertinent for their activity, all the more so when these traces are generated by the users themselves.

Systems Which Foster Reflexivity

Reflexivity, defined as the capacity to reflect upon one's activity, is deemed crucial in learning processes, be it for simple adjustments of one's actions or attitude in the course of one's activity (self-awareness) or for meta-cognitive regulation tasks that involve assessment or extension of one's knowledge, individually and collectively (Yukawa, 2006). Reflexivity can be fostered by systems that provide users with two kinds of elements upon which to build a reflective activity. The first level is related to *instantaneous reflection*: it provides *feedback on the activity* such as clues about the effectiveness of one's actions such as mirroring of the activity (Jermann et al., 2005), real time collaboration indicators (Nova et al., 2006), temporal/social representations of current group activity (Farooq et al., 2007) or peer feedback (Phielix et al., 2009). The second level of reflexivity involves *delayed reflection*: it provides *representations of the interaction beyond its scope* such as histories or Edit or Read ware (Hill et al., 1992), "reflective follow-up" representing one's way of solving a problem (Katz et al., 1992), indicators summarizing the interaction within a group (Bratitsis et al., 2006), or activity summaries (Laffey et al., 2009).

Trace-based Systems

The present research project focuses on *traces of interactions* between users and systems. Those that should not be confused with traces generally defined as what remains of an activity, including documents produced by participants. Logs are obviously traces for engineers who need to understand the functioning of the systems they build. While some approaches have tried to design log-mining techniques in order to obtain higher-level user activity-related traces (see Romero & Ventura, 2007 for a survey of the educational use of traces), our own approach pertains to research dealing with explicitly designed rich histories of interaction. Such traces can be used either by professional analysts to understand the activity that has taken place (Avouris et al., 2007) or by users willing to make sense of their own activity by consulting Web navigation histories (Weinreich et al., 2006) or even by replaying online interactions in which they took part (Plaisant et al., 1999).

Because most such tracing systems are designed in an *ad hoc* way to test particular research ideas related to dedicated activities, it is rarely possible to reuse these tools in subsequent projects. To go beyond such punctual approaches, we have proposed a model-based approach to traces, formally defining Trace-Based Systems (Settouti et al., 2009a) as systems built upon "modeled traces" organized in trace bases. A trace model is an ontology of the *obsels* (observed elements) that can be collected in one trace. Trace transformations in trace bases allow simplifying, merging or rewriting traces so as to obtain higher-level traces that present users with adequate representations of their activity. A formalization of our proposal can be found in (Settouti et al., 2009b).

The VISU Tool

VISU is a trace-based tool that features a number of different modeled traces for instantaneous and delayed reflection in the context of synchronous tutoring interaction.

Three Levels of Traces in VISU

Our videoconference tool provides three types of trace. Each corresponds to a distinct obsel type:

- *Interaction traces* are histories of users' actions automatically collected in real time from their interactions with the software. Events such as "launch video", "open image" or "begin a new activity" are considered. In Figure 1, the raw interaction trace of user1 has been filtered (transformation τ1), and the resulting trace is shared. In Figure 2, the vertical timeline features (a) and (b) that correspond to representations of obsels that have been shared by learners with their tutor while (c) is a representation of the "beginning of a new activity".
- *Marker-based traces* refer to text annotations that are written during the activity describing events such as "vocabulary" or "problem with microphone". Such annotations are similar to notes that are usually jotted down by a teacher during a learner's oral presentation so as to provide her with feedback. These traces can refer to the tutor's activity, the learner's or their joint activity. In figure 1, the marker-based trace has been built (transformation τ2) from the marking trace (which describes the activity of creating, deleting, modifying markers). On the vertical timeline shown in Figure 2, (d) and (e) are markers that have been set by the tutor.
- *Comment-based traces* correspond to text annotations that are added after the activity with obsels such as "effective dialogue between learners" or "nice piece of dialogue". A comment-based trace can be considered as a *post-hoc* manual transformation / rewriting (τ3) of traces that have been collected during the activity. On the vertical timeline shown in Figure 2, (g) is a comment-based trace.

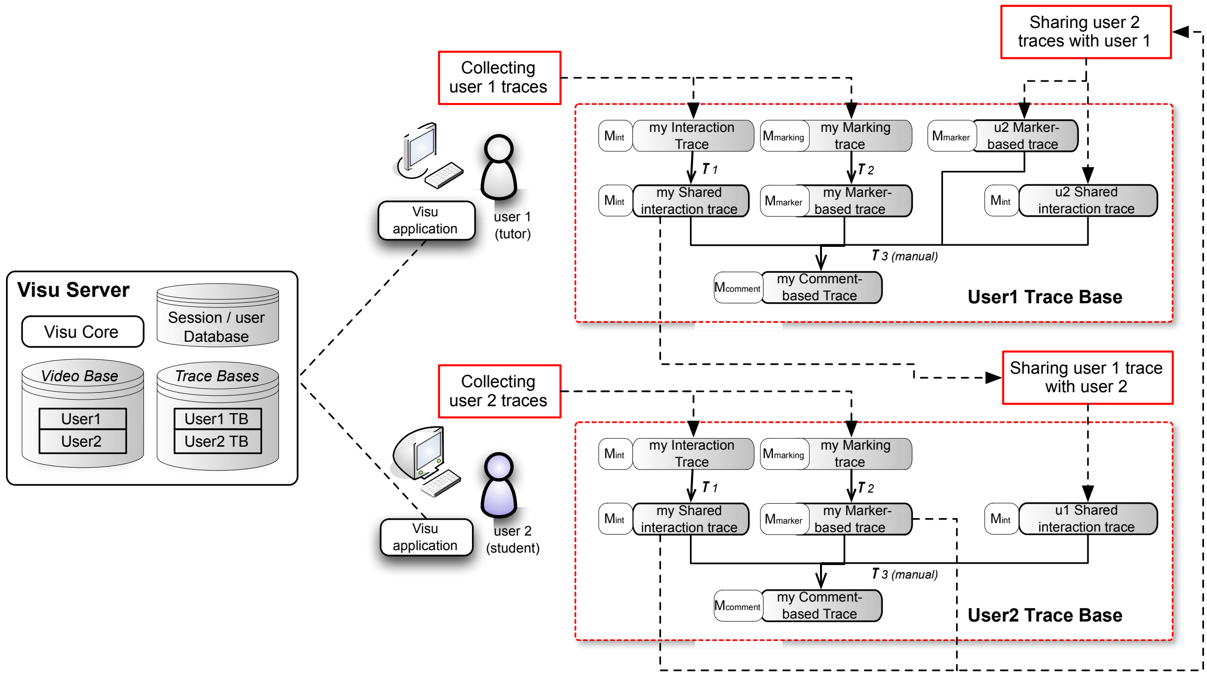


Figure 1. A tutor (*user1*) is interacting with a learner (*user2*). A video base stores the video streams that have been recorded during the interaction. Each user has her own trace base. *User1*'s trace base contains 7 traces, 5 of which have been collected from her activity, while 2 have been shared by *user2* from her trace base. *User2*'s Trace Base contains 6 traces, 5 of which have been collected from her activity, and 1 has been shared by *user1* from her Trace Base. As a tutor, *user1* does not share her marker-based trace while, as a learner, *user2* does.

Instant Reflection in the Interaction Room: A Synchronous Timeline for Activity Awareness and Activity Annotation

The interaction room (Figure 2 - left) is a classical videoconference room with a text chat. Tutors have access (on the left of the screen) to their “session manager”, which gathers planned session activities, instructions and keywords. The main feature which distinguishes VISU from other videoconference tools is the vertical “synchronous timeline”. This timeline presents users with selected elements from interaction traces (here “Image received” obsels represented as thumbnails) as well as marker-based traces.

VISU's synchronous timeline is designed to foster instant reflection by augmenting synchronous interaction with elements of other participants' activity (e.g. indicating when a document is accessed) and it allows participants to take notes on the activity with markers. Instant sharing of interaction traces is a means of extending collective awareness amongst participants. Using markers allows tutors to select critical events in the course of the activity in order to come back to them m in order (1) to provide learners with feedback at the end of the session and (2) to reflect on one's own activity in the retrospection room.

Delayed Reflection in the Retrospection Room: Obsels and Comments for Experience Reconstitution

The retrospection room (Figure 2 - right) basically functions as a video player allowing users to review previously recorded sessions (video and audio). The playback feature is supplemented by a horizontal timeline displaying the collected obsels (interaction + marker-based traces) organized either by obsel types or by users. The annotating tool (“*Mes commentaires*” [my comments]) allows the user to annotate events that occurred during the session. Such comment-based traces will serve as a basis for teacher development and be used for example in weekly trainee teacher debriefing sessions in which pedagogical interactions are analyzed.

The retrospection room in VISU is designed to foster delayed reflection. The delayed retrospection made possible in the retrospection room allows for teacher development in the following way. To learn professional competencies such as monitoring distant learner interactions effectively, providing clear and concise instructions, giving useful and timely feedback, trainees are confronted with traces of their own situated activity of a past online teaching session and are asked to select samples of this activity, analyze them and comment upon them during a collective debriefing session (Guichon, 2009). Several operations are required: noticing critical events that are worth commenting upon, categorizing them (what is the purpose of a specific sequence of actions?) and assessing their efficiency or lack thereof (was it adequate, timely, etc.?) against a professional repertoire (see for instance Hampel & Stickler, 2005). The aim is thus to help trainees to explore and reflect on a still ill-defined professional situation and retrospectively make sense of problematic events that occurred in the course of the activity. We contend that getting apprentices to analyze their own context-embedded practice is a necessary condition to make them aware of their behavior in context and a way to provide them with some kind of "annotated template[s] for future praxis in similar situations" (Olson, 1995: 121).

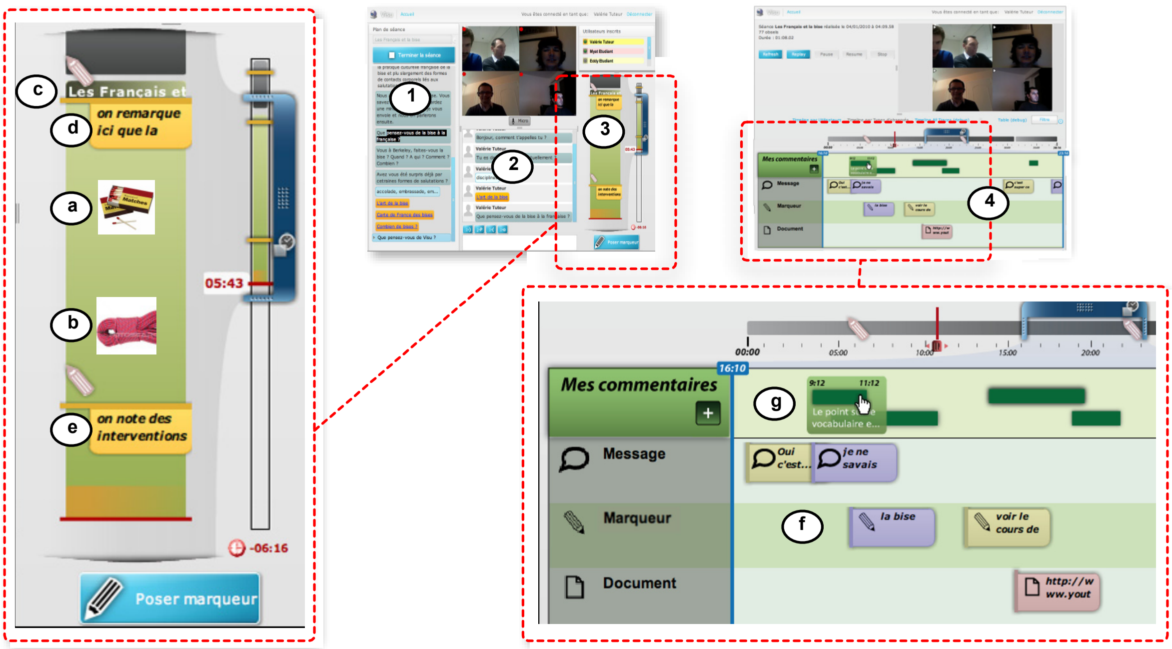


Figure 2. Interaction room (left) and retrospection room (right) in VISU1. The interaction room is used during the synchronous interaction: it provides (1) a session manager, (2) a chat window and (3) a vertical timeline, which features several presentations of obsels from both the interaction trace (a, b, c) and the marker trace (d, e). The retrospection room is used asynchronously after the interaction. It has a horizontal timeline (4) that presents markers and interaction obsels (f). It also allows the user to add comment obsels. (g) Unlike interaction trace and marker-based trace obsels, comment-based trace obsels can have a duration.

A Pilot Study of Marker-Based Traces

An exploratory study was conducted in real teaching context in order to (1) investigate the quantitative and qualitative use of markers during synchronous language teaching sessions and (2) assess the utility and usability of such a functionality for language tutoring. The ultimate aim was to further investigate design and training issues. Although the markers were also used in the delayed retrospective phase, only what happened in the interaction room was analysed. We focus here on the interaction trace, which appears on the timeline in real time and is also recorded in the chat history that remains available throughout the session.

Method – Experimental Data

The study was part of a telecollaborative project between the university of Berkeley and the University of Lyon 2. The VISU tool was used to support distant synchronous French lessons given by trainee French teachers in Lyon to language learners in Berkeley. Each session was organized in three phases:

- Each week, teacher trainees in Lyon (now referred to as tutors) had to prepare a lesson that aimed at fostering oral communication and to conduct it with one or two learners in VISU's interaction room. At the end of the session, tutors were instructed to conduct a rapid debriefing with the learners on linguistic issues with the help of the markers they had set during the interaction.
- After the interaction, the tutors had to analyze the learning session on the basis of the traces recorded in VISU's retrospection room.
- The following week, teaching interactions were then displayed and discussed in a plenary session with all the tutors and trainers in order to raise language teaching and tutoring issues (see Guichon, 2009 for the detail of the training program).

Tutors were given an introductory training session on the VISU platform, but the use of markers was not a focus. Thus we have the opportunity to study the tutors' intuitive use of these markers as they learned to use the technology to teach their distance language sessions.

Data Collection and Analysis

The data regarding marker setting and chat messaging were collected from 8 tutors over 7 tutoring sessions. At the end of the third session, each tutor was interviewed in order to comment on the markers he/she created during the session. They were also asked about their global perception of VISU. A quantitative analysis was conducted on the number of markers and text chat messages over time and across the sample of the 8 tutors. In addition, qualitative analyses were conducted on the type of markers produced by 2 specific tutors, selected before the first session on the basis of their prior experience: These two tutors had similar one-year experience in face-to-face language teaching but none in online tutoring. Interviews data were used to complement the quantitative analysis of markers.

Results

Given that the tutors were not familiar with the tool and rather new to online tutoring, it was expected that they would find the early sessions cognitively challenging. We predicted that markers set by the tutors would improve over time, both qualitatively and quantitatively, for two non exclusive reasons: first, the tutors gradually develop competence with the tool and the task so that cognitive resources would be sufficient for carrying out this dual task; second, their perception of the utility of setting markers to prepare for the debriefing and the next tutoring session would rapidly increase once the markers had been used for language teaching and reflexive analysis.

We present quantitative analyses of the number of markers and chat edited by all tutors (Figure 3), followed by a qualitative analysis of the type of markers set by two tutors.

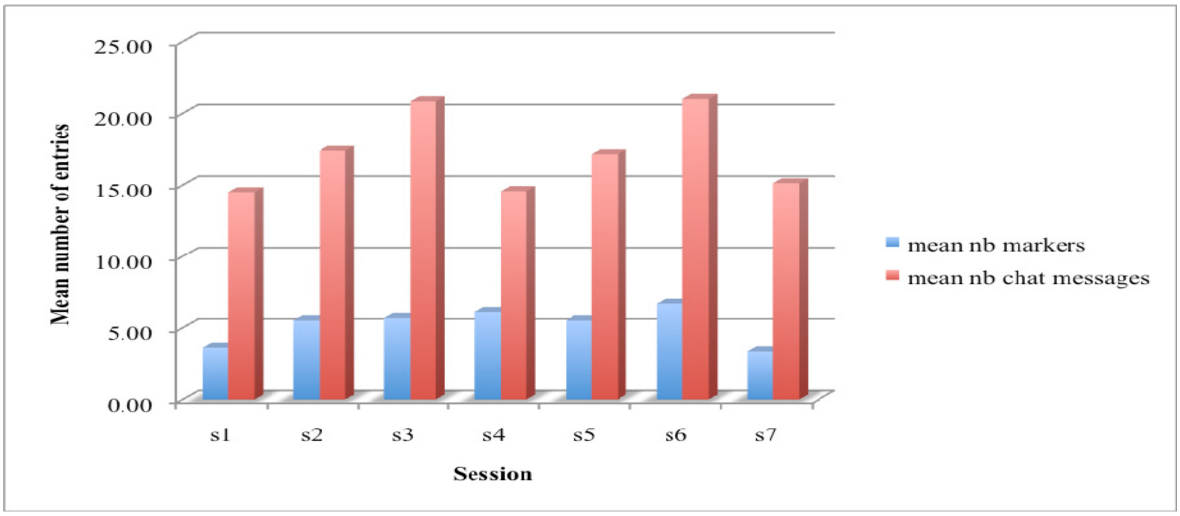


Figure 3. Mean Number of Markers and Chat Messages Averaged across Tutors over the 7 Sessions.

Number of Markers Edited over the 7 Sessions

As shown in Figure 3, the number of markers increased from the first to the second session and remained stable across the next sessions, except for a drop in the last session. Overall, the mean number of markers did not differ significantly across sessions ($F(6,42) = 1.624$, NS). However, a large variation across tutors was observed (min = 0.29 and max =12.7 mean number of markers per session) and the difference was statistically significant ($F(1,7) = 10.47$, $p < .05$). Moreover, there were significant correlations between the number of markers set by each tutor for session 2 to 6 (all $r > .80$, $p < .01$), but not between session 1 nor 7 and the others, meaning that in sessions 1 and 7 the tutors behaved differently than in other sessions. These results suggest that participants quickly became acquainted with the system and the task since they were able to set markers in a stable way as early as the second session. The drop in the number of markers set in the 7th session is probably related to the fact that it was the last session. If the markers were used not only to keep track of linguistic issues during the session but also to prepare for the next session, there would be less need to set markers in the last session.

Number of Chat Messages over Time

The tutors wrote a large number of messages in the chat (M = 15.94, SD = 12.56, see Figure 3). The mean number of messages seemed to vary over time but the difference did not reach statistical significance (F (6,24) = .343, NS). As for markers, the variation across tutors was large (min = 8 and max = 34.28 mean number of chat messages per session) and was statistically significant (F (1,7) = 28.73, p < .001). There was no correlation between the number of markers and the number of chat messages (r = .63, NS), suggesting that the variability across tutors was not linked to a general writing behavior but rather to a differentiated use of the marker and chat functionalities.

Qualitative Analysis of Markers

As mentioned in the methodology, the markers set by two tutors during their interaction with the distant learners across the 7 sessions were further analyzed in order to better evaluate the type and function of markers in this situation. We used Daele & Docq's (2002) typology, based on previous work on face-to-face and distant tutoring, which classifies tutor's interventions according to different purposes. It has to be noted that Daele & Docq's classification was elaborated to analyze tutor interactions with the learners while here the markers were also intended to regulate the tutors' own activity and were not displayed to learners. The four categories proposed by Daele & Docq (2002) were thus adapted as followed:

- *Social function*: pertaining to the quality of the communication and its social quality (e.g. "learner 1 is pleased")
- *Organizational function*: dealing with the organization of the session (e.g. "learner 3 arrives", "use of keywords")
- *Pedagogical function*: providing indications on learners' linguistic skills and their recurrent mistakes (e.g. "pronunciation of *fromage*")
- *Technical function*: corresponding to comments on the system (e.g. "learner 3 does not see the image")
- A fifth category (*other*) was created for markers that did not fit any of the above categories.

The classification (displayed in Table 1) was made thanks to the analysis of the content of the markers and the context in which they occur. In addition, the interviews with these two particular tutors were taken into account so as to better comprehend how they used the markers.

Table 1: Number of markers in each category for the two tutors (names have been changed).

Category	*Cindy*	*Nelly*
Social	13	0
Organizational	22	0
Pedagogical	37	48
Technical	8	0
Other	5	3
Total	*85*	*51*

As shown in Table 1, though pedagogical markers were the most frequent for both tutors, other specific uses of the markers can be observed. Nelly used them exclusively to take notes about pedagogical events that she could use for the wrap-up at the end of the session, or to prepare the next session. Cindy used the whole range of categories. Further examination showed that many of the markers created by Cindy were used to keep track of her evaluation of learner's behaviors, personality and skills ("S is not talkative" "they use the vocabulary from last session, good") and reflections on her own tutor practice ("+ good idea for the dialogue", "- I interrupted her").

An interesting phenomenon is the emergence of operative language over time. For example, "pronunciation" is shortened into "pron" by Nelly from session 3 on. Personal cues were sometimes used by Cindy, like + or –, to indicate when she evaluated the event as positive or negative, or capitals for emphasis ("SHE TALKS A LOT"). Interviews with the two tutors indicated the development of instrumentation over time. For example, Nelly first noted only "pronunciation" or the correct sentence without indicating where the mistake was. Then she found ways to keep track of what was said by the learner, for example by using capitals to indicate the place where the mistake was (e.g. "tellemenT"), or by noting a phonetic approximation of the learner's contribution rather than the correct sentence.

Usability and Subjective Utility

All the tutors were interviewed after session 3 about their perception of the tool and particularly their use of the markers. As expected, tutors appreciated the fact that they could set markers about learner's mistakes without interrupting the oral interaction. Moreover, they kept track of events they could use during the debriefing sessions with the other tutors and the teacher trainers. However, they found that the interaction was disturbed when they were in the process of setting markers because, thanks to the webcam, learners could see that the tutor was typing something that they could not see and was thus not looking at them at that precise moment. For these reasons, they reported using the markers for important events only and tried to find suitable moments (e.g., when the learner was looking at a video) to set markers. Regarding usability, they reported that setting a marker required more effort than writing in the text chat, particularly because the marker zone was very small. Finally they underscored that writing on the vertical timeline was not natural at all.

Discussion and Future Work

This study aimed at evaluating how novice language tutors used the functionality of setting markers during the synchronous session. Regarding marker setting, although markers were used overall as early as the second session, considerable differences in behaviour were observed across tutors, both qualitatively and quantitatively. Some tutors appreciated to keep track of learners' mistakes without interrupting the oral interaction. But some considered that the time taken and the attention required to set the marker caused a major disturbance in the flow of the activity and therefore they seldom used this functionality. Moreover, the qualitative in-depth analysis of two tutors' behaviors and perceptions revealed that the use of markers is not self-evident, one of the tutors exploiting all the possible functions of the marker while the other concentrated only on its corrective function. In addition, the tutors complained that the marker functionality was not easy to use. Therefore, design issues had to be examined since they dramatically affected the actual utility of the marker functionality, both as an external memory of the activity and support for practice. In this exploratory study, tutors have been explained how the marker functionality worked, but not how they could use it to improve the tutoring situation since no similar situation has been studied before. The present results suggest that tutors have also to be trained to make an optimal pedagogical use of markers, prompting tutors to use markers not only for keeping track of learners' mistakes, but also for fueling their reflection on their own practice and on the learning situation.

A second version of the application, VISU2, has been designed on the basis of the results obtained through this study. In order to avoid the split-attention effect and lessen the disruptive effect of setting markers, the marker text field has been enlarged and now appears next to the chat window and just below the webcam images. After being edited, markers are automatically moved to the right of the timeline. In addition, the timeline has been set horizontally to align it with the regular representation of time and ensure consistency with time representation in the retrospection room. Preliminary observations with the new version of VISU suggest that different tutors now set more markers than with the previous version of the tool. If confirmed, this result will reinforce the hypothesis that metacognitive or awareness tools need to be integrated into the workspace in order to be attended to without too much extraneous cognitive load (Buder, in press).

Future studies will focus on (1) evaluating how tutors make use of the markers in the retrospection room to analyze their tutoring activity retrospectively and (2) whether markers modify distant learners' behaviors and contribute to enhance their language competence.

Endnotes

(1) ITHACA (Interactive Traces for Human Awareness and Collaborative Annotation) is a project funded by the French Research Agency (ANR), see http://liris.cnrs.fr/ithaca

(2) VISU V1 adopts the architecture for trace-based synchronous collaborative systems defined in Clauzel et. al (2010). It is open source software (LGPL) based on Adobe® Flash/Flex and Red5 server. Olivier Aubert, Lionel Breduillieard and Serguei Sayfulin are the main developers.

References

Avouris, N. Fiotakis, G. Kahrimanis, G. Margaritis, M., & Komis, V. (2007). Beyond logging of fingertip actions: analysis of collaborative learning using multiple sources of data. Journal of Interactive Learning Research, 18(2), pp. 231-250.

Bratitsis, T., & Dimitracopoulou, A. (2006). Monitoring and analyzing group interactions in asynchronous discussions with the DIAS system. In *CRIWG 2006, Lecture Notes in Computer Science 4154*. 54-61.

Buder, J., Bodemer, D., Dehler, J., & Engelmann, T. (2009). SCAN tools for collaborative learning. In *International Conference of Computer Supported Collaborative Learning* (pp. 606–615), Rhodes, Greece.

Buder, J. (in press). Group awareness tools for learning : Current and future directions. *Computers in Human Behavior*, 10-13.

Clauzel, D., Sehaba, K., & Prié, Y. (2010). Enhancing synchronous collaboration by using interactive visualisation of modelled traces. *Simulation Modelling Practice and Theory*, In Press, Corrected Proof.

Daele, A., & Docq, F. (2002). Le tuteur en ligne, quelles conditions d'efficacité dans un dispositive d'apprentissage collaboratif à distance ? in *Proceedings of the AIPU Conference*, Louvain-la-Neuve.

Farooq, U., Carroll J.-M., & Ganoe, C.-H. (2007). Supporting creativity with awareness in distributed collaboration, in *International Conference on Supporting Group Work* (pp 31-40).
Guichon, N. (2009). Training future language teachers to develop online tutors' competence through reflective analysis. *ReCALL, 21*(2), 166-185.
Hampel, R., & Stickler, U. (2005). New skills for new classrooms: training tutors to teach languages online. *Computer Assisted Language Learning, 18*(4), 311–326.
Hill, W. Hollan, J. Wroblewski, D., & McCandless. T. (1992). Edit wear and read. In *Conference on Human Factors in Computing Systems (CHI'92)* (pp 3–9).
Jermann, P., Soller, A., & Muehlenbrock, M. (2005). From mirroring to guiding: a review of the state of the art technology for supporting collaborative learning. *International Journal of Artificial Intelligence in Education, 15*, 261–290.
Laffey, J., Amelung, C., & Goggins, S. (2009). A Context Awareness System for Online Learning: Design Based Research. *International Journal on E-Learning, 8(*3), 313-330.
Lonchamp, J. (2006). Supporting synchronous collaborative learning: a generic, multi-dimensional model, *International Journal of Computer Supported Collaborative Learning, 1*, 247–276.
Nova, N. Wehrle, T. ,Goslin, J., Bourquin, Y., & Dillenbourg, P. (2006). Collaboration in a multi-usergame: impacts of an awareness tool on mutual modeling. *Multimedia Tools and Applications, 32*(2),161–183, 2006.
Oviatt, S., Coulston, R., & Lunsford, R. (2004). When do we interact multimodally? Cognitive load and multimodal communication patterns. In *Proceedings of the 6th Int. Conference on Multimodal interfaces* (pp 129–136), Penn. State College.
Olson, M. R. (1995). Conceptualizing narrative authority: Implications for teacher education. *Teaching and Teacher Education, 11,* 119–135.
Phielix C., Prins F.-J., & Kirschner P.-A. (2009). The design of peer feedback and reflection tools in a {CSCL} environment, in *International Conference CSCL* (pp.626-635).
Plaisant, C., Rose, A., Rubloff, G., Salter, R. & Shneiderman, B. (1999). The design of history mechanisms and their use in collaborative educational simulations, in *Proceedings of the International Conference on Computer Supported Collaborative Learning* (pp. 348–359).
Romero, C. & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications, 33*(1), 135-146.
Settouti, L.-S., Prié, Y., Marty, J.-C., & Mille, A. (2009a). Trace-based system for technology-enhanced learning systems personalization, In *9th IEEE International Conference on Advanced Learning Technologies*, Riga, Latvia.
Settouti, L.-S., Prié, Y., Champin, P.-A. Marty, J.-C., & Mille, A. (2009b). *A Trace-Based Systems Framework: Models, Languages and Semantics*, LIRIS Research Report, inria-00363260, 40 pp.
Weinreich, H., Obendorf, H., Herder, E. & Mayer, M., (2006). Off the beaten tracks: Exploring three aspects of web navigation. In: World Wide Web Conference (pp. 133–142).
Yukawa, J. (2006). Co-reflection in online learning: Collaborative critical thinking as narrative. *International Journal of Computer-Supported Collaborative Learning, 1*(2), 203–228.

Acknowledgments

We are grateful to J. Lüthi who carried out the data collection for this study and D. Pries, R. Kern and D. Malinowski from UC Berkeley who made this experiment possible in Berkeley and N. Guichon and C. Develotte at the University of Lyon 2. We also wish to acknowledge Shona Whyte's insightful remarks on a previous version of this paper.

A Collaborative Platform Supporting Knowledge Sharing and Problem Solving for Teacher Professional Development

Thierry Condamines, Laboratoire MIS, Université de Picardie Jules Verne, 33 rue Saint Leu, 80039, Amiens Cedex 1, France, thierry.condamines@u-picardie.fr

Abstract: The present paper proposes a platform to support teacher professional development. From a study of exchanges between teachers on Web forums we examine two facets of knowledge exchange, usual teaching practices and problem solving, and how they can be managed on a unique web platform. We use a task/method paradigm for knowledge elicitation about usual teaching practices, driven by the exchanges between teachers and capitalized into individual memories. For problem solving we use an IBIS-like approach (problem, solution, argument) and REX-like cards to structure the problem and solution description. The important problem of knowledge evaluation is treated with a three-level evaluation (opinion, usefulness and difficulty). The teacher's profile and its central role in connecting teachers is described. For knowledge classification, because a consensus is difficult in this domain, we propose a mixed approach between structured classification techniques (taxonomies) and unstructured techniques (folksonomies).

Introduction

In France, like in many countries all around the world, teacher professional development is one of the main concerns of authorities in charge of Education, and the case of young teachers is the most critical one. The first years of their career are a moment where they oscillate between pleasure and suffering, facing a "double learning": that of their pupils and that of their job. If professional development has for a long time been considered as a linear process where teachers apply external knowledge developed by researchers, latest researches and learning theories (Knight, 2002; Butler et al., 2004) have shown that this process is continuous mixing formal training and informal training through discussions with colleagues and reflection on teaching practices (Uwamariya & Mukamurera, 2005).

The lack of formal training urges teachers to compensate by participating in Communities of Practice either at a local level (inside their school or institution) or at a global level on the Web gathering together people with common practices, interests and purposes (Wenger, 1998; Koh & Kim, 2004). There, teachers exchange ideas and experiences, help each other to develop their skills and expertise and exchange to solve very contextual problems. In this community the dynamics of the exchanges favor innovation (creation of new methods and practices) thanks to debates and practice confrontation (Daele, 2006) (see Figure 1).

On one hand, at the local level, teachers exchange in a face-to-face relationship without computer mediation and avoiding capitalization of the knowledge created. On the other hand, at the global level, Web technologies like forums, blogs and wikis allow the accumulation of exchanges but in an unstructured and poorly contextualized way. Thus, knowledge created through discussions is hardly reusable. Moreover, the number of online forums in France where teachers talk about their practices limits knowledge creation, each one reinventing what has been discussed in another forum. This is due to the fact that the environments used are not specific for CoPs as stated by numerous works. Tools are often designed either to support sociability to the detriment of contextualized resource production and their reification, like Tapped In (Schlager & Fusco, 2004) or CoPe_it! (Karacapilidis & Tzagarakis, 2007), or to support resource accumulation to the detriment of sociability, like in doceNet (Brito Mirian et al., 2006). New works try to unify the two approaches, in an effective application of Web 2.0 principle, like Cloudworks (Conole et al., 2008) and TE-Cap (Lavoué & George, 2010), by supporting exchanges between members of the CoP and between CoPs and also capitalizing on these exchanges by placing them in the associated context. In these tools, a message on a subject, which can contain educational resources, a question about a difficulty, or a simple idea, is posted by a member and shared by the CoP allowing other members to react by posting messages, sharing resources … But separation is blurred between teaching practice description and requests from a teacher seeking help about a very contextual difficulty in his own practice. As in online forums, practice description is often given by a teacher A as an answer to a request for help from a teacher B and is actually an adaptation of the practice in the context of the described problem (what A would do in the situation of B) and can be different from the real practice of A in his own context (which can also be interesting to capitalize).

In fact, when we observe formal training in teachers training institutes (called IUFM in France) and lifelong training programs we can see two facets of knowledge construction. First, teachers have to observe other (experienced) teachers in different teaching contexts which allow them, by reflecting on the observed practices, to construct their own practice. They can then experiment their new practice in their own context encountering difficulties and asking colleagues or experienced teachers for help. Following this observation, our work aims to provide CoPs of teachers with a unique Web platform which will support:

- knowledge elicitation (teaching practice) guided by the exchanges between teachers, capitalized into individual memories with the associated context;
- exchanges for problem solving, capitalizing problems (and their context) and solutions given by other teachers;
- Sociability and member participation with a Web 2.0 approach and a user profile management.

As a first stage in addressing this aim, we propose in the next part a description of the two facets in knowledge exchange and the models we use to take them into account.

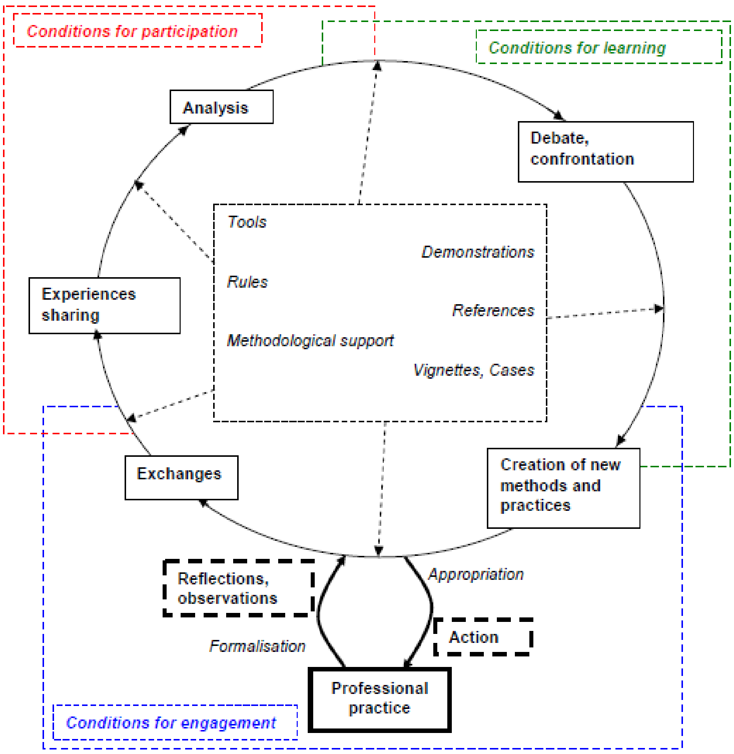


Figure 1. Daele Model of Teacher Professional Development through a CoP.

Two Facets in Knowledge Exchange

The first step of our work was to analyze online forums (http//forums-enseignants-du primaire.com is one of the most active), one of the tools mostly used in France by teachers to exchange on their practices, and cross the results with those observed in a teacher training institute. From this study we have classified the knowledge exchanged in two categories (see Table 1)

Table 1: Examples illustrating the two categories of the exchanged knowledge.

Usual teaching practices	Problem solving
"In your classroom, how do you proceed for diagnostic evaluations?" -------------------- *"This is a post for CE1/CE2-level teachers. Can youtell me what you do in this kind of two-level class and give me good ideas to improve my practice?"* -------------------- *"I need help to construct a time-table for a CE1-level class."* -------------------- *"Next year I will have a double level class. I have never had this configuration before. How can I organize my teaching (schedule, group work ...)?"*	*"I have a CP-level class with 24 pupils. 9 of them have big difficulties (2 of them having a MS level). I began with the book "L'école des Albums" but those 9 pupils don't follow [...] Do you have an idea and help me?"* -------------------- *"I have a pupil who seems to have no auditory memory. He can't neither remember the alphabet nor recite the first numbers list, [...] but he can [...]. What can I do for him?"* -------------------- *"I have a CM1-level class. We've worked on numbers until 999999 with all the class and at the same time on numbers until 99999 with 3 pupils having difficulties*

-------------------- *"I have a four-level class (MS/GS/CP/CE1) and I would like to share my experience on this kind of class. How do you make your time-table? How do you manage readers and non-readers…?"*	*and until 9999 with one having more difficulties. But now I don't know how to manage the next activities because I have introduced big numbers (million). What can I do for the 4 pupils still having big difficulties on smaller numbers?"*

The first category refers to what we call "usual teaching practices". This refers to tasks having to be achieved by all the teachers and the methods used to achieve them. It depends on the global context of the class (level) but it doesn't refer to a very particular situation. A teacher, often a young teacher, wants to know what others do in a given context (same class level). This reflects a need to share practices and to have reference teaching practices to analyze and from which to construct their own practice (Perrenoud, 1999).

The second category refers to a problem solving support. A teacher encounters a very specific problem depending on the global context of the class but also particularities like particular pupils (handicapped pupils, pupils with specific difficulties …). It can also refer to a particular activity constructed by the teacher, the use of a particular resource …

When a teacher wants to know what others do in their class (similar to his own class) in their day-to-day practice he is not interested in very specific problems. On the other hand when a teacher encounters a problem in his practice, this is always in a particular situation and he needs a specific answer. So our postulate is that these two facets can't be treated at the same level.

Usual Teaching Practices: The Task/Method Paradigm and an Exchange Driven Acquisition

In This Module Of our platform, we want to allow a teacher to construct a personal knowledge memory containing descriptions of his practices. In this process it's not only important to know what he does but also his reasoning. To represent the reasoning model (how a task can be performed), we used the Task/Method paradigm (Camilleri, 2003) where a task (what the teacher have to do) can be performed by different methods (a method describing one way of performing a task) giving a decomposition of the task into sub-tasks. A method represents the know-how which it's possible to implement in order to achieve the task. A method is characterized by conditions that must be satisfied to apply the method, controls giving the achievement order of the sub-tasks and a sub-task set. In this paradigm, a special type of tasks is available: terminal tasks. Terminal tasks are directly executable tasks requiring no decomposition (see Figure 2).

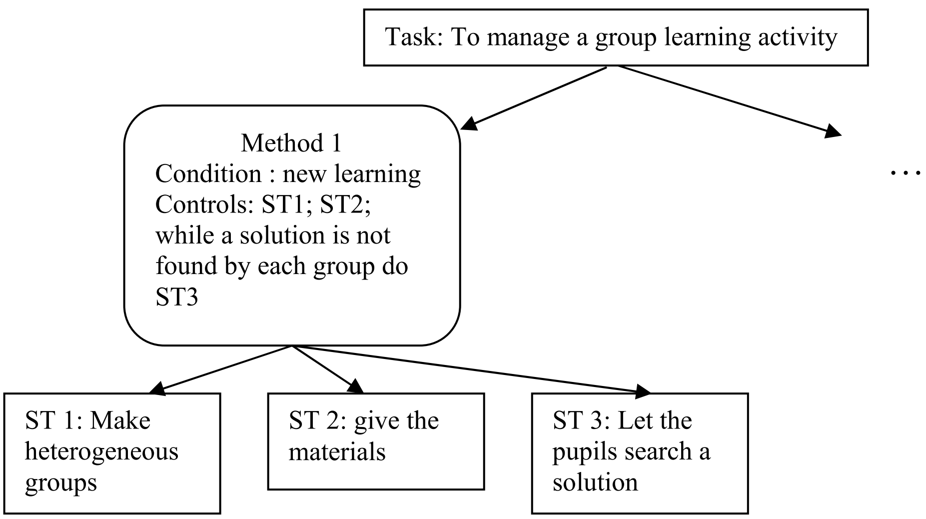


Figure 2. Example on Task/Method Decomposition (ST Means Subtask).

In knowledge management, knowledge acquisition is realized through discussions (interviews) between a knowledge engineer and experts. In our case if there are experienced teachers there is no representative expert. Each teacher have developed his own know-how which can be different from another teacher but equally effective. There is often no consensus on a teaching method. So our idea is based on knowledge acquisition from each teacher using the platform, storing it in a personal knowledge memory. Since interviews can't be done we use an exchange driven process. Starting from a reference list of basic tasks (high level tasks) each teacher have to achieve, the task/method decomposition can be started by a teacher wanting to share his practice and is refined through requests from other teachers (see Figure 3). It can be a request with a new condition getting to describe a new method. For example, in the case described on Figure 2, a teacher can ask "How do you manage a group learning activity in the case of reinvestment of an already learned knowledge?". A new method must be described with the new condition (reinvestment of a learned knowledge). It can also be a request for more details on a method or a terminal task. For example, a teacher can ask "how do you make the heterogeneous groups?". The method or the terminal task must be refined.

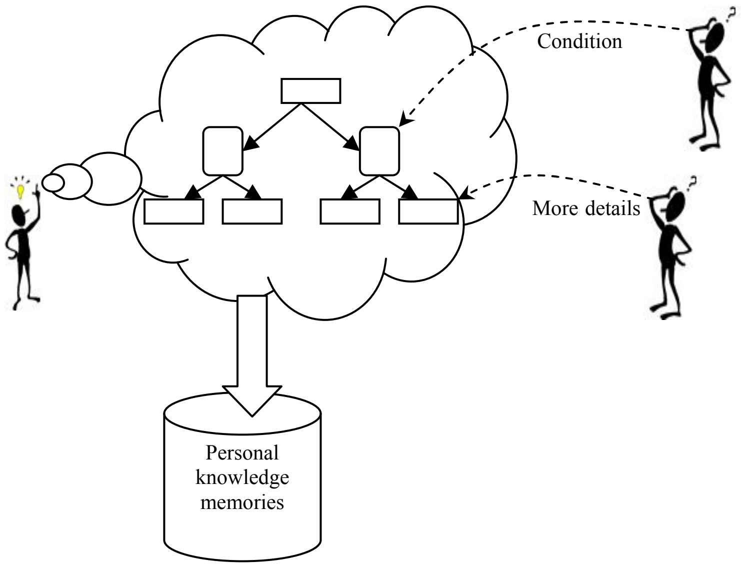


Figure 3. An Exchange Driven Knowledge Acquisition.

If a teacher ask for details, describing an unusual situation, it can't be managed in this part and enter in a problem solving process.

Problem Solving: The IBIS Approach

In this part, a problem refers to a difficulty encountered by a teacher facing an unusual situation in a very precise context and for which he needs help (see examples in Table 1). When we look at the answers on the forum we find different solutions given by different teachers but also arguments for or against a given solution. The exchange (problem, solutions and arguments) can also generate a new question referring to a new difficulty. For modeling those exchanges we use the IBIS method (Conklin et al., 1991) approach (which is traditionally used for capturing Design Rationale) with three elements: problem (defined by the context of the situation, particularly the task during which the problem occurs, and a question specifying the help needed), solution and argument.

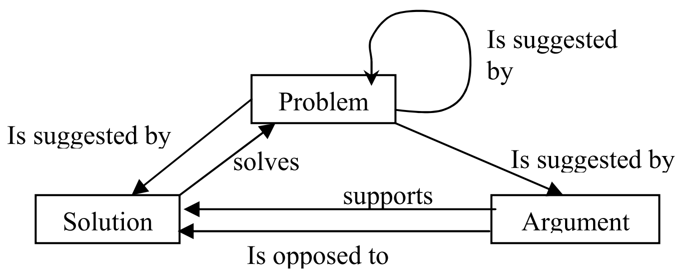


Figure 4. General Model of Problem Solving.

When a teacher describes a problem, he writes a problem card giving several characteristics:

- The task during which the problem was encountered. For example: "*Implementation of a teaching sequence – Collaborative working phase*". The first task level comes from the reference list of tasks mentioned in the previous part. Subtasks can be proposed (coming from subtasks already given for others problems or coming from the personal knowledge memories) and new subtasks can be entered.
- The type of the problem. For example: "*problem of discipline*". Types of problems are proposed coming from problems already defined and referring to the same task. A new type can also be defined.

- The context during which the problem occurs. It's different from the working context of the teacher which is defined in his profile as we will see later. For example: "*The children were very excited, speaking loudly, moving from one group to another. After one hour during which I was often shouting, I stopped the activity and ask for the solutions found. But there were no interesting results and I have to do it again later!*"
- The question that specifies the help needed. For example: "*If I do it again what can I do for more concentration of my pupils?*"
- Tags.

When another teacher wants to propose a solution, he can ask the problem author for more details. It could be done through a discussion between all the teachers like in online forums but it's long and difficult for a third teacher to read all the exchange and have a good idea of the problem and its solution. So we choose another way using a solution card and separate asynchronous discussions between the problem author and solution authors. The solution card contains a description of the solution, an arguments supporting it and recommendations to success in the solution practical use (especially for young teachers). This approach is inspired by the REX method (Malvache et al., 1993) used for knowledge capitalization and using knowledge card. When one ask for more details (problem author or solution author) the corresponding card is modified. The solution card is visible only by its author and the problem author until the description is clear for the problem author. When the discussion is closed, the solution card becomes visible to everyone and arguments can be given.

Knowledge Evaluation and Feedback

An important point is the knowledge evaluation process. Our choice was to use a three-level evaluation and rating:

- The opinion about a method (in the personal knowledge memory) or a problem solution which is important to identify communities having similar teaching methods or pedagogical approaches. A mark (from 0 to 4) can be given with an argument describing the reason of such a mark.
- The usefulness of a method or a problem solution. When a teacher try to use a method or a solution he gives a feedback with : a mark (from 0 to 4), an argument describing the reason of such a mark, a description about how he has applied and perhaps adapted the method/solution in his own class, and advices for a better use of the method/solution. This evaluation is very important, especially when it comes from young teachers.
- The difficulty to use the method/solution in the practice. A mark (from 0 to 4) is given when a teacher has used it and details about the difficult steps are given.

The usefulness and difficulty evaluations can be crossed with the teacher profile to know more about it for a given level of experience or class type.

Teacher Profile for Connecting People

The teacher profile is composed of:

- Identification data for accessing the platform (login, password, …)
- Pseudo: the fear to be judged is important in this community so anonymity is crucial.
- Age, sex
- Certifications
- Working context (see Figure 5): characteristics of the current class (level, number of pupils, standard or specialized class, presence of handicapped pupils), characteristics of the school (located in city center, suburbs, rural area, special locations like "violence area", number of classes), characteristics of the position (full-time, part-time, …)
- Past experiences: different working contexts
- Books used : this is an indicator about the pedagogical profile of the teacher since a book generally refers to a specific pedagogical method
- Centers of interest: tasks about which the teacher wants to be kept informed in priority
- Statistics about the teacher activity on the platform: number of connections, last connection, number of problems/solutions published, number of evaluations given …

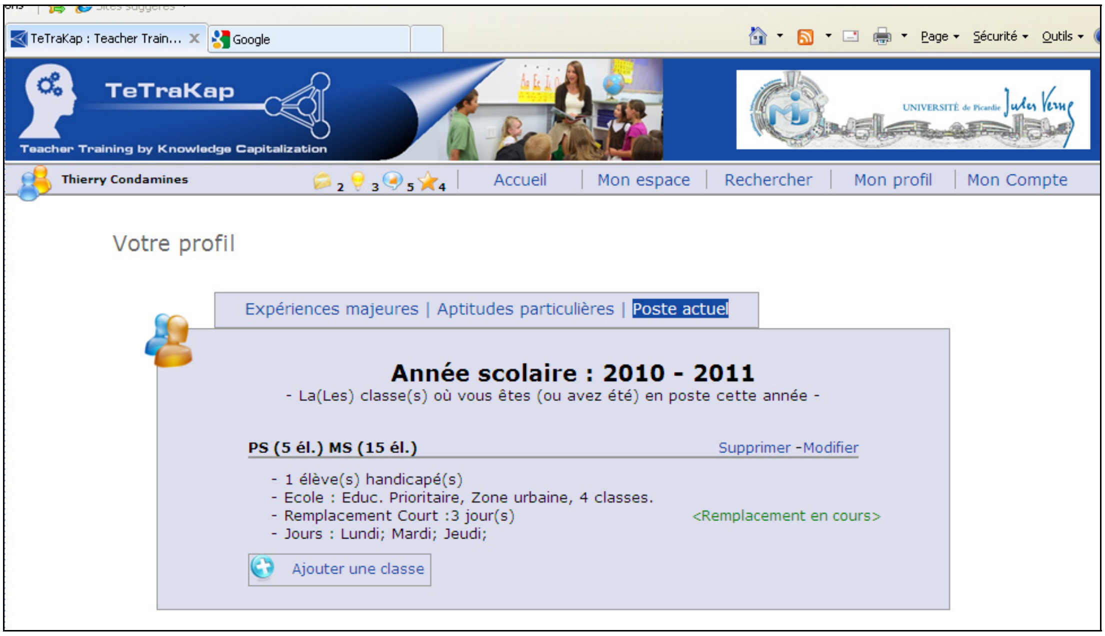


Figure 5. A Profile Element: Actual Working Context.

This profile is very important because it gives a better understanding of methods and problems but also to connect people. Knowledge broadcasting is a well-known problem and Van Heijst et al. (1996) gives different ways of collecting and broadcasting knowledge. When a problem is put in the system, its author generally needs help quite quickly. So to avoid passivity we choose an active broadcast (push way): when a problem card is written, it's published on the home page of teachers who are likely to give an appropriate help, that is people who have a sufficient experience in/near the same working context. In the same way, when solutions cards are published, the problem and its solution appear on the home page of teachers having quoted the task where the problem occurs as a center of interest.

Knowledge Classification

If taxonomies and ontologies are structured conceptual representations often used in knowledge indexing systems, they require a consensus between experts which is long and expensive to obtain and not always comprehensible by novices. When we observe the literature in the education field we can see that such a consensus is quite impossible. Moreover the evolution in time of these models is difficult to manage and is the subject of lots of recent researches. The other indexing system, the folksonomies, is less structuring but is based on the community activity and allows each user (knowledge as well as experts) to enrich the descriptions.

Our approach mixes the two systems, taxonomies and folksonomies. A general task classification (up to 2 levels) of high level tasks was constructed from official documents (Ministry of Education) on teachers' skills and interviews of teachers about their usual tasks. Subtasks can be added by each user through his personal memory construction or problem description. In addition users can add tags.

In addition, in a search, subjects are filtered by the teacher profile (working context) with a possible extension of the search space to other contexts.

Implementation and First Results

Our approach is implemented on a Web platform called TeTraKap (for Teacher Training by Knowledge capitalization) using PHP+MySQL technologies. The tests were realized with a group of 10 teachers from primary school with different level of experience (from novice to expert). Note that the technology was still under development during the time of study; therefore the results are mainly indicative.

The first criteria we used to know if the platform matches teachers' real needs were usability and sociability of the environment, level of knowledge sharing and adoption of the environment by users.

The community is of course too small to induce high interactions. But observing the activity on the platform and knowledge exchanged shows a progressive refinement of practice description (of experienced teachers) thanks to requests for details coming from novices. Moreover a novice can focus on an experienced teacher having a similar context and systematically asks him for help on specific problems (using the form for practice refinement). This effect shows that we must better think the separation between the two parts.

With interviews of teachers we could have details which can't be observed directly on the platform. If some tasks are not detailed because they are not crucial, even for young teachers, some of them are not because teachers need to exchange documents, timetables, photos of the classroom … this functionality is also important to help tacit knowledge elicitation, through videos for example. Concerning the interface, the "facebook-like" interface (see Figure 6) with profiles, practices and problems suggestions filtered according to the user profile seems to be appreciated. But the use of forms for practice elicitation is too heavy and not adapted. A graphical one could be a solution but this problem is still under study because the teachers don't seem to be familiar with some graphical tools presented.

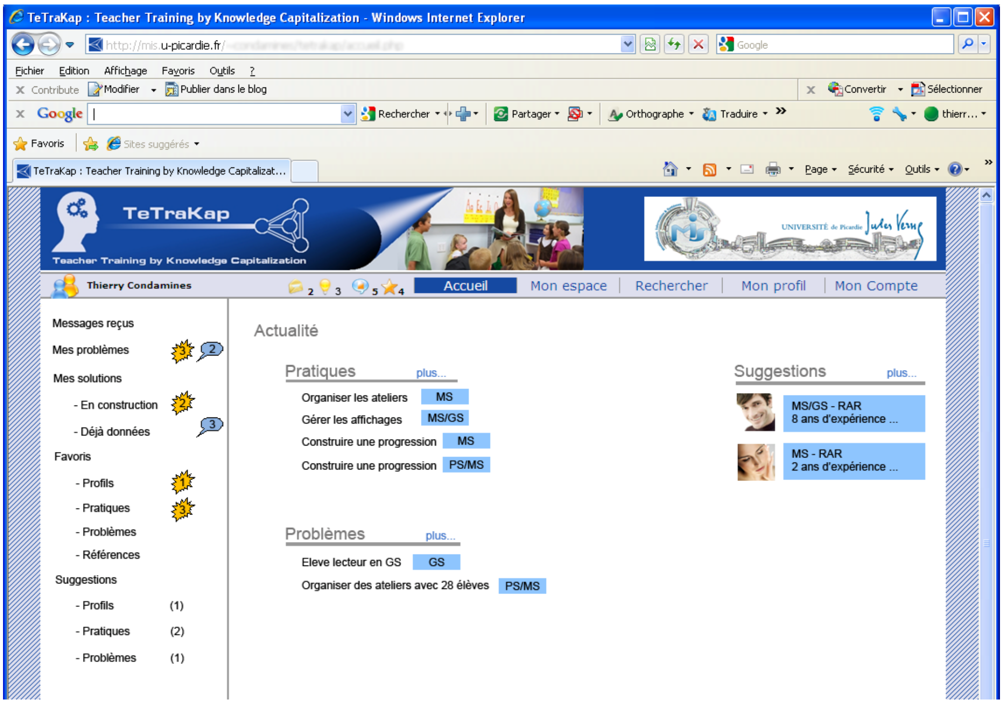


Figure 6. A Teacher Home Page.

Discussion and Further Work

Knowledge Co-construction

In this paper, we have shown a double approach of knowledge sharing for professional development in a community of practice linking up practice description and knowledge solving on a unique platform. This approach corresponds to a real and important request from teachers, especially young teachers or teachers changing context. Constructing your own practice by a try and error process is not only a long and difficult way for professional development but also a brake for innovation. In our approach, by connecting people with similar professional contexts, we will favor knowledge co-construction and innovation, allowing teachers to construct and improve their own practice by analyzing different existing practices. If the first results are encouraging the tests must be extended to a bigger population to evaluate the effectiveness of knowledge classification and the way people are connected, especially for problem solving. In this part, problems are indexed by the task/subtask where they occur and the problem type is given by the user. It would be interesting to use a high level problem taxonomy, as for the tasks, for better structuring. It would also be interesting to use a graphical interface, especially for the task/method decomposition of practices. But it's delicate because teachers, especially in primary school, are often reticent to use and construct computerized diagrams.

Bridging the Gap between Teachers' Practices and Educational Policies

A future work about our platform is to give functionalities to analyze the practices and problems. We believe that it could give indicators to adapt teachers training policies with better reactivity and an inventory of real teaching practices. On the other way it would be interesting to make it possible to education managers to give a knowledge "institutional evaluation" (but keeping anonymity, crucial for teacher participation). This 4th evaluation could give a return to teachers about the appropriateness of their practices to educational policies.

References

Brito Mirian, C.A., da Nobrega, G.M., de Oliveira, K.M. (2006). Integrating Instructional Material and Teaching Experience into a Teachers Collaborative Learning Environment. *First European Conference on Technology Enhanced Learning (ET-TEL 2006)*, pp. 458-463, Greece: Springer Berlin / Heidelberg.
Butler, D. L., Novak Lausher, H., Jarvis-Selinger, S. & Beckingham, B. (2004). Collaboration and self-regulation in teachers professional development. *Teaching and teacher education,*20, 435-455.
Camilleri, G., Soubie, J. L., Zaraté, P. (2005). Critical Situations for Decision Making : A support Based on a Modelling Tool. *Group Decision and Negociation*, 14, 159-171.
Conklin, J. & Burgess Yakemovic, K. C. (1991). A process-oriented approach to design rationale. *Human-Computer Interaction*, 6(3), 357-391.
Conole, G., Culver, J., Williams, P., Cross, S., Clark, P. & Brasher, A. (2008). Cloudworks: Social networking for learning design. *Ascilite 2008 conference*, pp. 187-196, Melbourne.
Daele, A. (2006). A model for representing professional development through the participation in a virtual CoP: uses for developing enhanced services. *Proceedings of the 1st International Workshop on Building Technology Enhanced Learning solutions for Communities of Practice (TEL-CoPs'06),* pp. 40-53, Greece.
Karacapilidis, N. & Tzagarakis, M. (2007). Web-based collaboration and decision making support: A multi-disciplinary approach. *International Journal of Web-Based Learning and Teaching Technologies*, 2(4), 12-23.
Knight, P. (2002). A systemic approach to professional development: Learning as practice. *Teaching and teacher education,* 18, 229-241.
Koh, J., & Kim, Y. (2004). Knowledge sharing in virtual communities: an e-business perspective. *Expert Systems with Applications,* 26(2), 155-166.
Lavoué, E. & Geaorge, S. (2010). Supporting the Interconnexion of Communities of Practice: The example of TE-Cap 2. *International Journal of Web-Based Learning and Teaching Technologies (IJWLTT),* 5(2), 37-57.
Malvache, P. & Prieur, P. (1993). Mastering Corporate Experience with the REX method, *Proceeding of the International Symposium on the Management of Industrial and Corporate Knowledge* (ISMICK'93), pp. 33-41. Compiegne, France.
Perrenoud, P. (1999). Du concret avant toute chose … ou comment faire réfléchir un enseignant qui veut agir. Retrieved November 5, 2010, http://www.unige.ch/fapse/SSE/teachers/perrenoud/php_main/php_1999/1999_20.html
Schlager, M. & Fusco, J. (2004). Teacher professional development, technology, and communities of practice: Are we putting the cart before the horse? In Barab, S., Kling, R. & Gray, J. (Eds.), *Designing for virtual communities in the service of learning*. (pp. 120-153). Cambridge : Cambridge University Press.
Uwamariya, A. & Mukamurera, J. (2005). Le concept de "développement professionnel" en enseignement: Approches théoriques. Revue des sciences de l'éducation, 31(1), 133-155.
Van Heijst, G., Van der Spek, R., Kruizinga, E. (1996). Organizing Corporate Memories. *Proceedings of the 10th Banff Knowledge Acquisition for Knowledge-Based Systems Workshop (KAW'96),* pp. 42-1/42-17. Banff, Canada.
Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge: Cambridge University Press.

Acknowledgments

We would like to acknowledge the teachers and colleagues for their participation in the study of the TeTraKap platform.

Interactive Representations of Student Activity to Inform Teacher Collaborations: Results from a Formative Exploration

R. Benjamin Shapiro, Morgridge Institute for Research, Madison, WI, ben@getdown.org
Peter Samuelson Wardrip, University of Pittsburgh, Pittsburgh, PA, psw9@pitt.edu

Abstract: We describe a student activity visualization tool that we constructed to provide a context for ethnomethodological inquiry -- a design-based breach experiment (Crabtree et al, 2004) -- into how teachers might think and communicate about student thinking when they have tools that give them better access to student activity. We use the Learning to Notice framework (van Es & Sherin, 2008) to characterize the kinds of activity that occurred when teachers used the tools in a team meeting. Then, we show how theses uses suggest ways that research on and uses of student activity representations may need to be sensitive to how contexts of implementation shape the consumption and use of those representations.

Objective

Researchers in CSCL, educational data mining, games for learning, CSCW, and HCI more broadly are working to develop manual and automatic techniques for characterizing and classifying patterns, including patterns over time, in users' collaborative and individual interactions with computer-based media (Reimann, 2009). Suthers and colleagues are developing a methodology, and supporting tools, for analyzing logs of online activity in order to discover patterns in uses of on-screen representations that span multiple participants (Suthers et al., 2010; Medina & Suthers, 2008). Shute et al. (2009) describes how the play of educational games may be scrutinized to create psychometrically valid assessments of learning and understanding that may be embedded in-game. Bit by bit, CSCL and its cognate fields are developing the methodological tools to produce precisely the kind of accounts of student activity that could support breakthroughs in the responsiveness of instructional environments (Fullan et al., 2006). But very few efforts have been made to actually construct systems that would assist *teachers* in understanding their students' activity and learning in CSCL environments. Rare exceptions include Feng and Heffernan (2006)'s ASSISTment project, which explores how students' interactions with an intelligent tutoring system may be summarized in order to help teachers to know what students know and Summary Street, which provides teachers and students with feedback about the quality of students' text comprehension and summary writing (Wade-Stein & Kintsch, 2004).

Because of this, very little progress has been made toward figuring out how practitioners (teachers or students) will use computer-generated accounts of student activity (and what the challenges to doing so will be) once the difficult methodological challenges of constructing them have been sufficiently overcome. We do not believe that research into the application of representations of student activity and learning to improving teaching must wait until methods for generating valid and reliable accounts of activity and learning are established. Instead, we propose that CSCL should deepen its inquiry into how representations of students' activity may be used in schools even as it figures out how to create those representations. Indeed, research on how such accounts could be used, including the diverse ways in which different teachers might interpret different representations or in which different backgrounds or classroom contexts might shape their interpretation may be quite valuable in shaping the learning analytic program by helping researchers within it to focus their efforts on techniques that might have the largest impact on practice.

This paper describes the beginnings of an attempt to do just that. We describe a student activity visualization that we constructed to provide a context for ethnomethodological inquiry -- a design-based breach experiment (Crabtree et al, 2004) -- into how teachers might think and communicate about student thinking when they have tools that give them better access to student activity. We describe our results in two parts: first, we use the Learning to Notice framework (van Es & Sherin, 2008) to establish whether our tool provides enough stimulus to the existing context of work to make activity occur that is consistent with a leading theory of teacher learning. Then, we show how the uses of our tool in teachers' collaborative practice suggest ways that research on student activity analysis representation construction and uses of new representations in practice may need to be sensitive to how contexts of implementation shape the consumption and use of those representations.

Theoretical Framework

Teaching is a socially complex dynamic of data-rich interactions between teachers, students, and content, mediated by classroom materials. What teachers choose to focus on with respect to their instruction, their classroom environment, and student understanding has a powerful influence on what they understand about their classrooms (Sherin & van Es, 2003). Teacher noticing is simultaneously a process of directing attention to certain features of instruction as well as a sense-making activity that connects what one notices to larger principles of action. Collaborative practices that improve teachers' capacity to notice student thinking in classroom activity can also develop common professional vision (Goodwin, 1992) amongst teams of teachers (an important ingredient in improving instruction in schools) as well as make teaching practices more public amongst faculties. In order to understand teacher noticing, Sherin & van Es have proposed a framework for Learning to Notice (2002). Within this framework, the skill of noticing consists of (van Es & Sherin, 2008):

- Identifying what is important in a teaching situation;
- Making connections between specific events and broader phenomena in teaching and learning.
- Using what one knows about the context to reason about a situation;

While the framework has been developed within the context of school-based video clubs (Sherin & Han, 2004; van Es & Sherin, 2008), research on teacher noticing has also developed with other techniques such as digital photography and teacher journal writing (Sherin & van Es, 2003). Ultimately, as teachers develop their ability to notice, they are able to reframe their discussions about instructional matters in terms of student thinking (Sherin & Han, 2004). Cross-culturally, the capacity of groups of teachers with diverse perspectives to draw attention to, and make varying interpretations of, a broad range of classroom events is core fixture of Japanese Lesson Study (Lewis, 2000; Fernandez & Yoshida, 2004).

We propose that the learning to notice framework could be a benchmark that the CSCL community uses to gauge the utility of different representations that it constructs about students' learning activities. If an implementation of a new tool does not yield practice that satisfies the Learning to Notice criteria, then some measure of redesign may be necessary before more in-depth analyses of practice are warranted. If, on the other hand, the Learning to Notice criteria are satisfied, then the implementation should offer researchers to opportunity to examine deeply the ways in which the new tools are used to construct meaning about student learning.

To explore this claim, we use the Learning to Notice framework to characterize the kinds of activity that result from the introduction of a data visualization tool that depicts students' literate activities into teachers' work. Specifically, we will show how the activity that resulted from the visualization tool's implementation satisfied the Learning to Notice criteria, then illustrate how particularities of the resulting activity illustrate complexities in teachers' reasoning about student thinking that should guide ongoing design-based research into representational tools for teachers

Context

This design-based research is part of a larger effort to address literacy in the content areas at middle and high schools. The essence of the larger work is to build effective reading-to-learn environments for students and to use these environments to help us understand the reciprocal relationship between content area achievement and reading achievement.

This work took place in an urban suburb of a large Midwestern American city in a school containing K-8 students. The school attracts a diverse group of students from both disadvantaged and extremely affluent families. In particular, this paper will focus on one group of sixth grade teachers over the course of the first year of technical assistance involvement. After listening to a presentation on literacy in the content areas for the whole district, this team of teachers self-selected for participation in the project, and their principal agreed to set aside paid, protected time (free of other obligations) for the teachers to participate.

In this group of teachers, Ms. N. teaches mathematics, Mr. C. teaches reading and language arts and Ms. T. teaches social studies and science. They teach three blocks of classes every day, which reach all of the sixth grade students. Each class is taught every day except for science and social studies, which are taught on alternating days. Reading and Language Arts also alternated, although they often blurred together as one block. In addition, up to three researchers facilitated the meetings, Mr. B., Mr. S. and Mr. W. The group met every Friday morning for approximately one hour.

We focus on building reading-to-learn environments that rely on the three strategic approaches to reading support: summarizing, T-Charts and annotating text (Scherer et al., 2008). Summarizing allows students to capture the gist of a chosen text in writing as well as the major concepts and details supporting those concepts. T-charts, also known as double-entry journals, are two or three column charts (like the shape of a T) that provide a structure for students to monitor and document their understanding of texts (Atwell, 1990). Annotation is the process of marking up a text in order to perform content analysis as well as reveal the meaning behind various textual features (Liu, 1996). While these techniques can improve student understanding, no research has examined how the representations that students produce through the application of these techniques can aid teacher insight into student thinking. The design based research described here was inspired by observations of teacher practice, which revealed that raw student annotation data was too voluminous and complex for teachers to use to drive instructional decision-making. Computer-based data visualization tools could help, but what should they look like? What kinds of teacher practices should they support? The absence of prior practice (both at our implementation site and in the research literature) offers no easy answers, and so we looked to a breach experiment, described in the next section, to find out.

Methods and Data Sources

We conducted a design-based breach experiment (Crabtree, 2004; Garfinkel, 1967) to illuminate the ways in which a team of teachers might use activity representation tools in their practice, and the personal perceptions, beliefs, and expectancies that shape those uses. Breach experiments are violations of the expectations that characterize normal social order, through the overt need they create for participants to restore that normal social order. This induces participants to explain how they think about normally tacit aspects of their lives, including expectations for others, and how the breach transforms or violates those expectations. The point is not to frustrate participants, but to violate their expectations enough that they must actively negotiate with their surroundings to restore normalcy. This is a valuable ethnomethodological tool because it may offer insights into participants' desires, perceptions, and dependencies that are so deep, and so tied to the participant's cultural milieau, that they difficult to interrogate through questioning and observation alone. They are, to use Garfinkel's language, "demonstrations designed as 'aids to a sluggish imagination'…produc[ing} reflections through which the strangeness of an obstinately familiar world can be detected." Crabtree (2004) argues that the introduction of novel technologies to participants' lives can function as breaching experiments, provoking practices and revealing participants' beliefs that could not otherwise be seen and making them available for design reasoning.

The purpose of this design-based breach experiment is to explore the ability of a re-representation of students' literate activity to support teacher noticing and thinking about students' literate thinking in content areas, as well as to use the differing ways in which different teachers react to what the new tools show (or what they see the tools as showing) about their students to illustrate (at least a corner of) the space of reactive possibility to novel representational tools in teachers' practice.

Breaching Tool: Markup

To do so, we created Markup, a prototype tool for students to annotate texts and for teachers to examine students' annotations. While annotation can be a powerful tool for student learning, it is difficult for teachers to use annotations to understand student thinking because the collected annotations (which we conceptualize as external representations of students' thinking) that result from a homework assignment typically contain thousands of data points spread across dozens to hundreds of pages. To assist teachers, Markup's teacher user interface provided three different tools for viewing student work; the most used and discussed representation was the interactive heatmap (see Figure 1). The heatmap was an interactive heatmap, inspired by Hill et al.'s (1992) Edit Wear and Read Wear. The heatmap, as depicted in Figure 1, is simply a complete original text with the text's background shaded to indicate the frequency with which the text there was annotated. Regions that were annotated frequently (number of students who annotated / number of students who did the assignment) glowed red (the more frequent, the bolder the red). A teacher-user could click anywhere on the text to see a table of what each student wrote about the clicked text; this is depicted in Figure 2.

Teachers agreed to use Markup in their classes for about 6 weeks, during which time they reviewed students' annotations in Markup before school on every morning that followed an evening when students were assigned reading with annotation for homework. They also used the tools within a weekly team meeting to present, review, and discuss reading comprehension in each others' classes. We observed the ways in which teachers used the tools to exploratively analyze students' work and the ways that they used the tools to empirically ground discussion in their team meetings.

Because we focus here on teachers' use of the tool in collaboration with each other (and not the work teachers did on their own), data for this paper are drawn from field notes and recordings from the team meetings. The specific data here come from two different team meetings, though many more were recorded and analyzed. These conversations were typical. As participant observers, we gathered observation notes, meeting agendas, and teacher-generated artifacts as well; this heightened our sensitivity to context of the research (Marshall & Rossman, 2006). These data provide thick qualitative descriptions of the teachers' individual and collaborative work. Using the Learning to Notice framework, the data were coded, often fitting into more than one of the Learning to Notice framework categories. The coded data also provided us with counterexamples. Analytic memos were written to test our conjectures and counterexamples during the analysis process (Maxwell, 1996). These memos were also an opportunity for discovery, orienting and developing new categories of analysis (Strauss, 1987). Two researchers discussed the data in order to increase reliability of the coding, multiple coding and counterexamples.

Results

In the results, we will use excerpts of teacher conversations during team meetings – selected for their illustrativeness – to illustrate how resulting teacher practice corresponded to the dimensions identified by the Learning to Notice framework. These examples indicate that our design-based breach led to new teacher practices. Moreover, these new activities correspond to those that in other interventions have tended to improve teaching outcomes. That is, the results illustrate how new tools may support a richer teacher focus on student thinking. In our discussion, we discuss how the particulars of *how* teachers did so suggest implications for further research into representational tools for teachers.

The boy indicated that he understood very well.

The next day the old man left the house after warning the boy to take care to do exactly what he had been told. During the afternoon, the boy put the beans on the fire to cook. Then he was filled with curiosity. What was behind the little door he had been forbidden to open?

Without any fear the boy opened the door and discovered in the room three enormous covered water jars. Then he found three capes inside a large trunk. There was one green cape, one yellow cape, and one red cape. Not satisfied with these discoveries, the boy took the top off the first water jar to see what it contained.

Immediately the water jar began to emit great clouds that quickly hid the sky. Frightened and shivering with cold, the boy opened the trunk and put on the red cape. At that instant a clap of thunder exploded in the house. The boy was turned into thunder and lifted to the sky, where he unleashed a great storm.

Figure 1. The Heatmap.

[illegible]	Character Says	He didn't even say hello; He only asked for food.
[illegible]	Character Says	He wants something to eat.
[illegible]	Character Says	not kind
[illegible]	Character Says	He makes a statement.
[illegible]	Character Says	he is poor or greedy

Figure 2. Pop-over Aggregating Students' Interpretations of a Text Passage Clicked by the Teacher.

Identifying What is Important in a Teaching Situation

Analyzing the transcripts of the group of teachers talking about the student work represented in the heatmap, we notice that the teachers are drawn to the features highlighted in the heatmap. That is, a common point the teachers call out concerns the extent to which students did or did not highlight the text. For example, Mr. C, the literature teacher, stated, "I am already really concerned that, like – why didn't – why did so few of them even do questions? I mean, that was the whole – that was the instruction. I mean, is – how many do you think this is?" He is referring to the disappointing amount of students as a whole class that actually did their homework. It is not surprising that teachers' attention would be drawn to this fact since this is the clearest function of the heatmap. As we note in our discussion, however, the fact that the heatmap encourages this way of looking at student activity may actually distract from the overall aim of improving teachers' understandings of student thinking.

The heatmap enabled the teachers to focus on individual students as well. While this also led to some observations of how much or how little annotation was done, the teachers also identified places in the text that were not annotated by individual students. For example, Mr. C observed about a student's annotation that "It is interesting, a ton in the beginning and then only a couple at the end… But the middle of the story does kind of get left out." Since the teachers had annotated their respective texts before assigning them to their students, they had an inclination of important places in the text.

The pop-up window depicting what questions or comments students made about their annotations also provided details that teachers called out. For example, when reading a question from one student, the science teacher says,

> **Ms. T**: Look at this. *[She reads what the text says:]* "Some stars are much larger than the sun" *She then reads a student's annotation.* "Dose that mean they are hotter? (sic.)" It's such a simple question but it really is important.

Here the teacher is calling out the question this student writes regarding a statement from the textbook. The student is connecting size of a star with heat. This is an important relationship within the scientific domain the students were studying; the tool supported the teacher noticing students' reasoning about it.

Making Connections between Specific Events and Broader Principles of Teaching and Learning

We found limited evidence of teachers explicitly linking specific instances of annotations to broader principles of teaching and learning. Nonetheless, several examples illustrated how teachers' *implicit* values shaped practice with Markup. One example worth noting refers to an exchange among the teachers about a student's question from the text:

> **Ms**. **T**: For example… [the text] says, "The sun is about 100 times greater than that of earth." They are trying to say how big the sun is and other stars and Daniel's [a pseudonym] comment is, "What about gas to keep the plane going if the plane were actually going that far?" … I mean he is thinking about this analogy.
> **Mr. C**: He's seeing outside the box.
> **Ms. T**: No. That's not even what I mean. I just mean he's actually like reading the words and understanding the words and this is maybe even what I'm talking about when I'm saying that sometimes it's so hard for me to say, "Well, a student doesn't understand this just by reading what they wrote," because there are so many ways –
> **Ms. N**: To access their knowledge?
> **Ms. T**: A student could have copied directly the definition for parallax from the text and I wouldn't say that I thought they necessarily understood it. I know so much more from Daniel saying, "What about gas to keep the plane going?" than I do from someone copying word for word the definition for parallax.

This excerpt demonstrates one teacher's reasoning about the particular student's response and how his response demonstrates understanding. Ms. T acknowledges that the student is trying to connect the size of the sun with an analogy from his own life. She is identifying the student's learning as a process of negotiation and interpretation (Cobb, 1994), including implications not intended by the author. She is not trying to determine whether the student's response is right or wrong (as Mr. C generally tended to), but rather whether the response reveals the kind of engaged attentiveness (an important broader teaching principle; see Pianta & Allen, 2008) to the text that the teacher is hoping for.

Using What One Knows about the Context to Reason about a Situation

Teachers' conversations about students' work also addressed student understanding. Ms. T stated that "It is harder to tell what they do not understand than what they do understand, because…I guess you can kind of say you are giving the benefit of the doubt, unless – like, I would want to see how they talk about it…" Here the teacher makes two important points. First, even with the new tools, she has a difficult time inferring what students do not understand about the text. Second, even when she infers that there is student understanding, she would prefer to check the understanding through another means, such as discussion. This presents an interesting problem: it is possible that annotation, which has previously been shown to *aid* text comprehension, may, alone, produce representations that cannot be easily used to *assess* comprehension. This suggests that one possible source of design guidance on designing representational tools for teachers, studies of existing assessment practices, may be of limited value to CSCL researchers in this area, because they do not encompass uses of data that can inform teachers (such as evidence of student engagement) but do not comply with psychometrically reliable models of assessment.

The extent to which the tool reveals or does not reveal student understanding was echoed later in the teachers' conversation. In a conversation with Mr. C, Ms. T said,

> "I just think …the ways that they…show that they understand it, like, you do not know exactly what those will be. Like I have to wait and see, how did they word something, or like what – how did they respond to it? And just because they, like, copied the definition directly from the book about light year [I might not know]"

Here the teacher is referring to the feature that allows students to write a definition of a word. While a student may be able to write the definition (definition was a kind of annotation; teachers sometimes directed students to annotate new vocabulary thusly) on the computer, it does not mean that they will be able to use the word in an appropriate way. The teacher is also reporting on the difficulty of articulating in advance what evidence of understanding will look like, and that instead a Stewart-like "know it when I see it" approach is needed. Mr. C responded by suggesting that Ms. T alter the task she has given students to require them to use the language in a sentence, rather than simply defining it. His suggestion illustrates the important role that the instructional context can have on the utility of any data visualization.

Another way that reasoning about context can occur is in using the insight into student thinking offered by a tool to make decisions about altering the context, such as by changing instructional plans based upon what one has seen. An instance of this occurred when Mr. C made note of students annotating in ways that were contrary to his thinking. This provided a tension in his mind between students not annotating "correctly" either because they were not annotating what he felt they should or because he was gaining evidence of multiple perspectives from the students. This provided new awareness – and instructional ramifications -- for the English teacher. As he told the other teachers, "if I discussed this story today I'm going to be going in more open minded that there are a lot more opinions out there than I thought there were."

Single Example: Learning to Notice

In some cases, an extended comment or discussion exhibited the multiple dimensions of the Learning to Notice framework. For example:

> **Mr. C**: …Dominic was done with it, got his ten done and stopped, didn't even get half way through the story. Whether he read it or not he was done. That is Dominic.
> **Ms. N**: That's so him.
> **Mr. C**: And that's his big issue, it's about as quick as I can get it done. The rule was ten, he did ten. So, it's hard to assess him. When you look at him, what he did in the beginning was – it was okay. He used a lot of the same words over, but he didn't do a horrible job, but that's him, half way. Get it done, walk away.

Here, within the context of Mr. C attempting to relate his instructional goals to what the student actually did, Mr. C is identifying what one student has done in the class and how the extent of the student's work is not enough to fulfill the full purpose of the assignment or to allow him to assess the student's understanding. This exchange corresponds to all three elements of the Learning to Notice framework because Mr. C called out a notable element of a student performance, characterized it in terms of a known problematic phenomenon (not completing an assignment), and related it to that student's historical tendencies. The conversation continued:

> **Ms. T**: I'm sorry, what about that thing he wrote for rap?
> **Ms. N**: He was really intense about that.
> **Ms. T**: He wrote the longest rap and it is awesome. It's so sophisticated.
> **Ms. N**: But it's the first time I've seen him like –
> **Mr. C**: Engage.

Here, Ms. T argued with Mr. C and Ms. N's historical characterization of Dominic by presenting a different background to the students' performance on the Markup. While Dominic's annotation may have been typical of his behavior in reading, his level of engagement had seen peaks (corresponding to an assignment that aligned with the student's interest in rap) as well. While this conversation does not necessarily address instructional strategies to address Dominic's engagement, Ms. T's gentle critique does allow the teachers to acknowledge the many sides students bring to their learning situations, including the ways in which well-constructed assignments can motivate students to excel. It is also noteworthy to mention that this snippet of discussion is not very deep with respect to the instructional topic. However, it is possible to view the collaborative actions among teachers through the Learning to Notice framework both to provide future support to the learners as well as to aid in the design process of future instructional tools by suggesting conversational contexts and protocols (e.g., presenting counterexamples that explain students' successes contextually) in which they might be employed.

Significance

This study demonstrates how design-based breach experiments can fill an information gap that exists when trying to design tools to support learning in an absence of current practice. This is acutely the case for researchers designing data analysis tools for teachers to use to understand student thinking through detailed analyses of student activity. With the exception of early mathematics and reading, such practices are exceedingly rare in schools and have been only lightly researched. We demonstrate how a breach experiment may be used to understand some of the practical problems of understanding students' literate activity through analysis of external representations produced during that activity. In addition, we showed how the Learning to Notice framework provides designers of collaborative learning tools a window into how teachers attend to student learning through examination of student activity. Looking at and discussing student work is an important learning activity for teachers and the learning is embedded in the collaborative activity (Koschmann, 1996). By providing a reference frame for identifying teacher practices that are associated with improved teaching, Learning to Notice allows CSCL researchers to identify whether an important class of teacher activity results from the implementation of a new tool, and then to study and potentially problematize that activity. As we showed, it is possible for individual teachers to practice in ways that satisfy the Learning to Notice criteria without thinking deeply about student thinking. For example, the Markup heatmap directed teachers' attention to how (in)frequently students annotated any given passage of text. But while this is an important feature, it challenges designers to highlight features that qualitatively differentiate student work and expose student thinking. Mr. C's past practice (i.e., before this intervention) tended to be strongly top-down; he directed classroom conversations based upon his interests (early field notes noted the lack of opportunities for students to talk about their ideas in his classroom) and he often judged students' activity, as depicted in the heatmap, based upon how well it concurred with his own. The heatmap supported this behavior by enabling him to make quick work of checking whether students annotated the "right" things. If used solely on his own, Mr. C's work with the tool may not, thus, have been much of a learning experience for him. The tool made it easy for him to continue his past practices and relative lack of focus on student thinking. However, with critique from his peer teachers, he recognized that other interpretations of the data were possible. By making activity in Mr. C's class transparent to his peers, Markup enabled a collaboration that expanded Mr. C's thinking about his students' activity.

Cases like these indicate ways that CSCL researchers could examine the problems of practice that occur when teachers try to use data analysis tools to inform their teaching. Specifically, by examining differential interpretations by different participants of common representations, researchers may understand not only the possible approaches to meaning making with new tools, but also the ways in which collaborative practices may bridge different interpretations to enrich teacher thinking. Such cases can aid design work by suggesting ways in which designed representations could either foreground details in student activity that are likely to have controversial interpretations (thus increasing the likelihood that teachers' conversations will include discussions of multiple points of view) or directly suggest multiple meanings – gathered through exhaustive study of how different teachers make sense of analogous cases -- for displayed data, tutoring teachers in how to make sense of classroom activity.

More generally, this work suggests visualization tools, even relatively simple ones like the Markup tool, can facilitate teacher learning and discussion through de-privatizing their practice. Extensive research has shown both the degree to which the teaching profession is isolated and the power of collaborative teams and communities of teachers to support learning. Prior research has suggested that a focus on a long-term instructional goal, such as literacy strategies, can aid in building collegial relationships among teachers (Wardrip, 2009). Our results show how even uncomplicated representations of student activity can support teacher analysis of student thinking, reaffirming our opening claim that research on responsive practices to new representations of learning activities need not wait until more complex analysis are ready for deployment.

In closing, it is important to note how these results highlight a limitation of the Learning to Notice framework with respect to studying collaborative learning. While the Learning to Notice framework highlights both individual and group-level change in perspective taking and noticing, it does not highlight tensions or dissonant views, including claims about appropriate instructional responses, that often occur in groups of teachers with varying experiences, beliefs and expertise. In future work, this theoretical framework may need to be modified in order to support deeper analyses of collaborative learning. CSCL has a pivotal role to play in guiding, and being enriched by, that theory construction.

References

Atwell, N. (1990). *Coming to know: Writing to learn in the intermediate grades*. Portsmouth, NH: Heinemann.

Crabtree, A., Benford, S., Greenhalgh, C., Tennent, P., Chalmers, M. & Brown, B. (2004). Supporting ethnographic studies of ubiquitous computing in the wild. ACM Symposium on Designing Interactive Systems, State College, PA, 60-69.

Cobb, P. (1994). Where is the mind? Constructivist and sociocultural perspectives on mathematical development. *Educational Researcher*, 23 (7), pp. 13-20.

Feng, M., Heffernan, N.T. & Koedinger, K.R. (2006) Addressing the testing challenges with a web-based e-assessment system that tutors as it assesses. Paper presented at WWW 2006, May 23–26, Edinburgh, Scotland.

Fernandez, C. & Yoshida, M. (2004) *Lesson Study: A Japanese approach to improving mathematics teaching and learning*. NJ: Erlbaum.

Fullan, M., Hill, P. & Crevola, C. (2006). *Breakthrough*. Thousand Oaks, CA: Corwin Press.

Goodwin C. (1994). Professional vision. *American Anthropologist*. 96, 606-33

Hill, W., Hollan, J., Wroblewski, D., & McCandless, T. (1992). Edit wear and read wear. Proceedings of the SIGCHI conference on Human factors in computing systems , 3-9. 201

Koschmann, T. (1996). Paradigm shifts and instructional technology. In T. Koschmann (Ed.), *CSCL: Theory and practice of an emerging paradigm* (pp. 1-23). Mahwah, NJ: Lawrence Erlbaum.

Lewis, C. (2000) Lesson Study: The Core of Japanese Professional Development. Paper presented at the special interest group on research in mathematics in education at American Educational Research Association Meeting (AERA), New Orleans, LA.

Liu, K. (1996). Annotation as an index to critical writing. *Urban Education*, 41: 192-207

Marshall, C., & Rossman, G.B. (2006). *Designing qualitative research*. Thousand Oaks, CA: Sage.

Maxwell, J.A. (1996) *Qualitative research design: An integrated approach*. Thousand Oaks,CA: Sage

Medina, R., Suthers, D. & Vatrapu, R. (2008). Inscriptions becoming representations. Proceedings of the 9th international conference on Computer Supported Collaborative Learning, Volume 1.

Pianta, R.C. & Allen, J.P. (2008). Building capacity for positive youth development in secondary school classrooms: Changing teachers' interactions with student. In M. Shinn & h. Yoshikawa (Eds.) *Toward positive youth development: Transforming schools and community programs*. New York, NY: Oxford University Press.

Reimann, P. (2009) Time is precious: Variable- and event-centred approaches to process analysis in CSCL research. *IJCSCL,* 4 (3), 239-257.

Scherer, J., Gomez, K., Herman, P., Gomez, L., White, J., & Williams, A. (2008). Literacy infusion in a high school environmental science curriculum. In K. Bruna & K. Gomez (Eds.), *Talking science, writing science: The work of language in multicultural classrooms*. NJ: Erlbaum.

Shapiro, R.B. & Wardrip, P.S. (2010) Understanding formative instruction by design. In: Proceedings of Ninth International Conference of the Learning Sciences, Chicago, IL, pp. 316-317

Sherin, M. G. & van Es, E. A. (2003). A new lens on teaching: Learning to Notice. *Mathematics Teaching in the Middle School*, 9(2), 92-95

Sherin, M. G., & Han, S. (2004). Teacher learning in the context of a video club. *Teaching and Teacher Education*, 20,163-183

Shute, V. J., Ventura, M., Bauer, M. I., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. In U. Ritterfeld, M. Cody, & P. Vorderer (Eds.), *Serious games: Mechanisms and effects* (pp. 295-321). Mahwah, NJ: Routledge.

Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences*. Cambridge: Cambridge University Press.

Strauss, A.L. (1987). *Qualitative analysis for social scientists*. Cambridge: Cambridge University Press

Suthers, D.D., Dwyer, N., Medina, R. & Vatrapu, R. (2010). A framework for conceptualizing, representing and analyzing distributed interaction. *IJCSCL,* 5(1), 5-42.

van Es, E. A., & Sherin, M. G. (2002). Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *Journal of Technology and Teacher Education*, 10(4), 571-596.

van Es, E. A. & Sherin, M. G. (2008). Mathematics teachers "Learning to Notice" in the context of a video club. *Teaching and Teacher Education*, 24, 244-276

Wade-Stein, D., and Kintsch, E. (2004). Summary street: Interactive computer support for writing. *Cognition and Instruction*, (22), 333-362.

Wardrip, P. S. (2009). The role of literacy work circles in developing professional community. Paper presented at the annual meeting of AERA, San Diego, California, April 13–17.

Unfolding Experienced Teachers' Pedagogical Practices in Technology-enhanced Collaborative Learning

Minna Lakkala, Liisa Ilomäki, Institute of Behavioural Sciences, P.O. Box 9, 00014 University of Helsinki, Finland
Email: minna.lakkala@helsinki.fi, liisa.ilomaki@helsinki.fi

Abstract: Several studies address the problem that the pedagogically meaningful usage of technology does not propagate easily; new methods for teacher training are needed. One method is to distribute experienced teachers' cultivated practices as examples, but good ways of presenting the essential aspects of these exemplary practices should be developed. The study examined how two experienced primary teachers constructed technology-enhanced collaborative learning units in order to document these examples as training materials. The first case was an inquiry project in biology; the second case focused on the virtual publishing of a Web journal. The Pedagogical Infrastructure Framework, including technical, social, epistemological and cognitive components, was applied in the analysis. The results revealed that experienced teachers planned long-term, goal-oriented processes consisting of well-organized activities where various elements of the pedagogical infrastructure were coherently integrated, which provided pupils with the possibility to practice challenging technical and academic skills in a meaningful context. Based on the analysis, ideas for the further development of teacher training materials are discussed.

Introduction

Modern digital technologies are considered to provide new, valuable possibilities for education. Yet research results indicate that pedagogical change in schools through information and communication technology (ICT) has not actualized as expected. Technology is still infrequently used by few teachers and is often relegated to traditional teaching methods such as sharing information or doing simple, mechanical exercises instead of being used to advance collaborative work, creative activities and the solving of complex, authentic problems or to support students' knowledge management skills (OECD, 2010; Smeets, 2005). ICT is not used to transform pedagogical practices but to support the teaching of domain content, and the school use of ICT is boring and monotonous (Gibson & Oberg, 2004; Pedersen et al., 2006).

In teacher training programs, technology, pedagogy and subject domain content are often taught separately, which does not necessarily provide teachers with competencies for integrating them adequately into classroom practice (Hyo-Jeong & Kim, 2009). Teachers would benefit from concrete, practical examples of using technology in real classroom settings and the combining of technical and pedagogical training (Pedersen et al., 2006; Putnam & Borko, 2000). To develop exemplary material for in-service teacher training, we examined and documented how two experienced teachers applied digital technology in their real classroom practices. The main idea was to disseminate examples that represent challenging pedagogical approaches, including the rich use of technology and complex activities, and to draw on experienced teachers' well-tried and cultivated practices. Many studies have examined the innovative usage of ICT in teaching (e.g., Kozma & Anderson, 2002), but the studies have not focused on unfolding experienced teachers' established, refined classroom practices with ICT in detail, in order to create best practice models for other teachers.

The aim of the present study was to investigate how two experienced primary school teachers constructed a learning setting for their pupils through the overall design and timely scaffolding of classroom activities. The Pedagogical Infrastructure Framework, developed in our previous studies, was used as an analysis framework for unfolding and describing the pedagogical elements and solutions in the expert teachers' designs and practices. First, the Pedagogical Infrastructure Framework is shortly explained, after which the methods and results of the study are described. At the end of the article, some conclusions are made about the issues that appear exceptional and worth disseminating in experienced teachers' pedagogical practices with technology.

The Pedagogical Infrastructure Framework

To unfold the central elements in the teachers' pedagogical practices, especially in technology-enhanced collaborative knowledge creation activities, we illustrated the organization of an educational setting as building up an appropriate *pedagogical infrastructure*. The idea for using the notion of infrastructure as a metaphor was adopted from previous studies (Bielaczyc, 2006; Paavola, Lipponen & Hakkarainen, 2002), to emphasize how the pedagogical design of collaborative settings is indirect in that it sets up background conditions that mediate intended social and cultural practices but does not prescribe the contents or tasks in detail (Jones, Dirckinck-Holmfeld & Lindström, 2006).

The approach has been developed further in our own studies (e.g., Lakkala, Muukkonen, Paavola & Hakkarainen, 2008), resulting in the *Pedagogical Infrastructure Framework*. It specifies four essential support structures - *technical*, *social*, *epistemic* and *cognitive* – that are considered important to include in the design of learning settings. The components are chosen in order to highlight aspects that are essential particularly for promoting technology-enhanced collaborative knowledge creation practices, but which are not necessarily emphasized or systematically considered in conventional pedagogical practices. The framework, as such, is not normative; it merely helps to examine and evaluate the design features in a structured fashion, drawing the designers' or researchers' attention to some critical but often unnoticed elements in the setting (Lakkala, 2010). The basic components are defined in the following way:

1) *Technical conditions*: What kind of technology and tools are in use? The provision of technology and technical advice to the participants, the organization and orchestration of the use of technology, and the functionality of the tools provided and their appropriateness for the desired activity.
2) *Social conditions*: How is collaboration organized and supported? The combination of designed individual or collaborative student activities and required outcomes; the organization of students' collaboration and social interaction.
3) *Epistemological conditions*: Why, how and by whom is knowledge produced? The ways of operating with knowledge and knowledge practices that the assignments promote, the nature of the information resources used, and the role of the participants and information resources in working with knowledge.
4) *Cognitive conditions*: How can the intended activity and competence improvement be explicitly supported and scaffolded? Designed tasks and artifacts or tools that perform a modeling and reflective function for promoting students' self-regulative competencies to work in an intended way.

A teacher's role in setting up technology-enhanced collaborative learning settings can be divided into two main tasks (see also Hogan & Pressley, 1997): first, planning, organizing and structuring the overall activity by establishing the underlying pedagogical conditions for the collective effort (*overall design*), and secondly, participating in the activities as a guide and expert who provides adapted, situation-specific guidance to the learners when needed (*scaffolding*). In the present study, both the overall design and the scaffolding activities of the experienced teachers were examined through the components of the Pedagogical Infrastructure Framework.

Methods

The general methodological approach in the present study was a multiple case study approach (Yin, 2003) where two primary school teachers' educational units in their classrooms were investigated. For the data analysis, we used descriptive and qualitative content analysis methods to unfold the teachers' pedagogical practices based on multiple types of data.

Research Context

The educational cases investigated in the study represented the Finnish cases in an international FICTUP project (Fostering ICT Usage in Pedagogical practices; see http://www.fictup-project.eu) supported by the EU. The aim of the project was to create training materials (written pedagogical scenarios and related short videos) describing concrete pedagogical activities using ICT, to experiment with a close tutoring process among experienced and novice teachers, and to provide training institutions with recommendations based on the experimentations. The present study focuses on the pedagogical units carried out by two experienced teachers from Finnish elementary schools. The training materials created for the cases are available on the FICTUP project website (FICTUP pedagogical scenario, 2010a; FICTUP pedagogical scenario, 2010b).

Participating Teachers and Their Pedagogical Cases

Two primary school teachers participated in the study, conducting technology-enhanced collaborative learning units in their classrooms. The teachers were experienced in the pedagogical use of ICT, they had participated in various development projects, and they were acknowledged teacher trainers on the communal level. Their practices with ICT have also been evaluated as advanced in previous studies (see the case 'Rome' in Lakkala, Lallimo & Hakkarainen, 2005, and the case 'Do you eat healthily' in Ilomäki, Lakkala & Paavola, 2006), and they were chosen as tutors for the FICTUP project because of their expertise in the pedagogical use of ICT. The cases were designed and conducted by the teachers themselves, and they represented the teachers' ordinary practices, which were not earlier influenced by the Pedagogical Infrastructure Framework.

In *Case 1*, titled 'Exploring growth factors', primary school pupils from the 3rd grade (9-10 years of age) studied wild courtyard plants and practiced the construction of a simple experimental design related to growth factors. They carried out an inquiry project in small groups, practicing scientific skills such as formulating research questions and hypotheses, searching for information, making and documenting observations, and writing and commenting on scientific explanations. The duration of the unit was ten lessons over three weeks; successive lessons varied from one to three lessons during the same day. Online working spaces of the Web-based collaboration system Fronter (http://com.fronter.info/) organized by the teacher were used to structure the inquiry process. The educational objectives of the unit related to learning the subject domain content (to understand the role of growth factors for wild courtyard plants) as well as to improving more general skills, such as writing, information search, categorization of knowledge, science skills, collaboration skills and technical competence. The inquiry process had various phases where the students took turns working alone (when producing and commenting on research questions and explanations), in pairs (when searching for information from literary sources) and in teams of four (when examining the courtyard plants, explaining observations and presenting the outcomes). Varying tools of the web-based collaboration system (a brainstorming tool, discussion forums, co-editable documents etc.) as well as an interactive whiteboard (IVB) were used for documenting and sharing ideas and contributions, depending on the nature and requirements of the activity in each phase. *Teacher 1* conducting the unit was a male teacher with 20 years of teaching experience, 19 of them at the current school. He had used ICT for 20 years in teaching, 24 years altogether. He was the teacher in charge of ICT issues at his school. He had been an active developer of ICT in education also outside the school for several years, working as a teacher trainer and scriptwriter of educational materials.

In *Case 2*, titled 'School children's Web journal', 5th grade primary school pupils (11-12 years of age) created stories for a digital school journal. About one third of the pupils had special education needs. The project lasted for six weeks, with two lessons at school each week, twelve lessons in all. The journal was published with the Magazine Factory program (http://www.mazinefactory.edu.fi), which is a free publishing application for schools that simulates professional publication practices. Besides that, the Microsoft Notepad application, digital cameras and IVBs were used. The aims of the unit were to become acquainted with virtual publishing and editorial tasks as well as to develop skills for creating digital material through story writing, taking pictures with a digital camera and pasting pictures and text in the publishing tool. The pupils were given the responsibility for creating stories in pairs or groups. They decided the themes, wrote the stories, took digital pictures and made proposals for the layout. Moreover, they commented on each others' drafts in the middle of the process and together evaluated all the outcomes and the working process at the end of the unit. *Teacher 2* conducting the unit had 26 years of teaching experience, two of them at the current school. He had used ICT for 15 years in teaching, about 17 years in all. He was an active developer also outside the school and had belonged to an ICT expert teacher team in his city for several years, training other teachers to use various ICT tools in teaching.

Data Collection
All the lessons constituting the pedagogical units conducted by the two teachers were observed and videotaped by one or two researchers. The teachers were shortly interviewed before and after each session about that particular lesson. In the pre-interview, the questions were about the goals, plans and expectations concerning the lesson and the usage of technology. In the post-interview, the teachers were asked whether the goals and plans were achieved and what they would do differently next time. All the outcomes produced by the teachers and the pupils during the process in the virtual working spaces and the Web were collected. The training materials for the FICTUP project (the written pedagogical scenario and related videos) were co-constructed together by the teachers and the researchers, based on the collected data and draft plans written by the teachers.

Data Analysis
The qualitative data analysis method used in the study was based on the approach introduced by Chi (1997). The primary data were the videotaped lesson observations because we wanted to focus on the teachers' actual practices in the classroom. All the videotapes were examined by the principal investigator of the study, and those sections revealing a design solution or scaffolding activity of the teacher were categorized qualitatively using the components of the Pedagogical Infrastructure Framework into whether the activity focussed on technical, social, epistemological or cognitive support. The detailed video data analysis was conducted using the Atlas.ti program. The main distinctive phases in the pupils' working process in each unit were defined based on the classroom observations and the written scenario. The distribution of the pedagogical infrastructure elements visible in the practices was then counted for each phase in order to get an overview of the teachers' pedagogical designs. Other data were used as a complementary source for interpreting the classroom activities and providing examples of the practices and reflections of the teachers.

Results
The results are presented separately for each case, and in both cases the ways that teachers designed and scaffolded the pupils' learning efforts in the units are unfolded. First, the distinctive phases of the process are shortly described, then the distribution of the pedagogical infrastructure elements in each phase is presented with a diagram. Sophisticated examples of the teachers' pedagogical practices in technology-enhanced collaborative learning are presented in a table, structured according to the four central pedagogical infrastructure elements. Finally, the main issues from the teachers' self-reflections after the lessons are summarized.

Pedagogical Practices in Case 1: Exploring Growth Factors through Inquiry

In Case 1, the teacher structured the working process in phases that for the most part followed the Progressive Inquiry model, which had been adopted by the teacher several years before in a previous development project (Lakkala et al., 2005). The teacher also presented the model to the pupils in a visual form (see FICTUP Pedagogical scenario, 2010a). The pupils' concrete activities in each lesson were shaped by these phases. The following phases of the process were identified:

1. *Creating the context*, including orientation to the topic, introduction of the progressive inquiry model and demonstration of the working phases in the Web-based collaboration system.
2. *Generating questions and forming the research groups*, where pupils generated questions about growth factors with a brainstorming tool. The questions were examined together using the IVB, and four research groups were formed based on them (light, water, warmth and soil/nutrients).
3. *Writing explanations of the research questions,* where pupils wrote explanations that answered their group's research question in a discussion forum and commented on other group's explanations.
4. *Making and documenting observations*, including field work where pupils measured the length of plants in two areas in the school courtyard. The results were written down in the pupils' notebooks.
5. *Interpreting and explaining observations*, where groups constructed a table about their observations in the groups' Web document, compared the length of plants in two areas and wrote interpretations of the differences. Then the pupils commented on the explanations of the other groups in a forum.
6. *Adding explanations from information sources*, where the pupils first in pairs sought new information about their research topic from a text source on the Web and then added relevant explanations to their group's Web document as a group.
7. *Giving presentations*, where each group presented their results in front of the class by explaining the content of their Web document (topic, results, interpretations and new knowledge found) with the IVB.

In all, 184 separate sections representing the pedagogical design solution or scaffolding activity of Teacher 1 were chosen from the 10 lessons of Case 1. The sections were distributed in the four support categories in the following way: 23% technical, 25% social, 35% epistemological and 17% cognitive. Figure 1 presents the distribution of these activities into categories in each phase of the process.

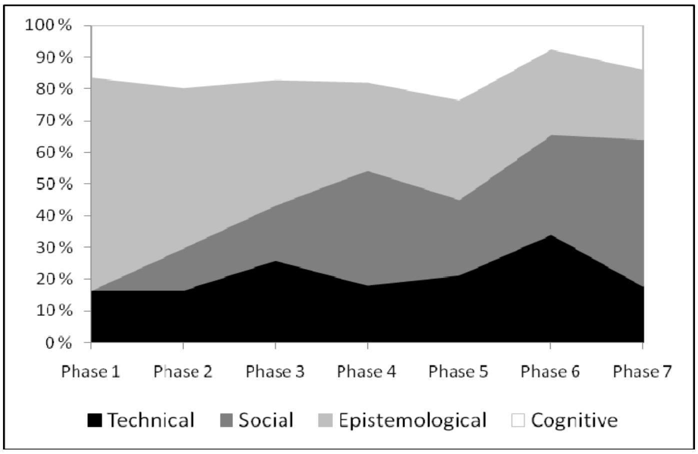


Figure 1. The Relative Distribution of the Pedagogical Infrastructure Elements in the Phases of Case 1.

Figure 1 shows how the focus of the designed activity in the first and second phases of the process was more on epistemological issues concerning the topic of inquiry and the research questions. Social issues received more emphasis from the teacher when the pupils started their group work or shifted between working individually, in pairs or in groups. What is noteworthy is that the roles of all four supporting elements were quite evenly distributed throughout the process. In Table 1 we have chosen examples of pedagogical practices by Teacher 1, which can be regarded to represent sophisticated, expert-like practices that are good models for other teachers. Separate examples are presented for overall design solutions and for scaffolding activities. When a practice relates to a specific phase of the process, the phase is mentioned in parentheses.

According to the post-interviews after the sessions, the teacher was generally satisfied with his design of the unit and the students' engagement. He did not do much evaluation of whether the process was successful or not, but carefully and critically analyzed which practices were challenging for pupils and in which skills they still required more guidance and practice (e.g., asking questions, making a table, understanding written instructions, sharing responsibilities in group work). The teacher thought that the same design could well be applied to various age groups and topics, if the participants' competence level is taken into account.

Table 1: Examples of sophisticated pedagogical practices in Case 1.

Support	Overall design	Scaffolding
Technical	The introduction of versatile tools relevant for the specific activity in different phases (brainstorming tool, discourse forums, Web document, IVB etc.).	Giving advice about advanced practices for using various, appropriate tools for different purposes and for managing multiple working spaces.
Social	Versatile combinations of individual, pair and group work according to the process phase and nature of the activity. Explicit demands for groups to combine individual results and make a common report page and presentation.	Training good collaboration practices and reminding groups to follow them throughout the process. Constantly taking care that pupils take turns in tasks carried out in groups with shared resources (e.g., one computer).
Epistemo-logical	A relatively complex, overarching inquiry task simulating scientific work. Groups were formed on the basis of the pupils' own inquiry questions. (P2) The use of various information sources: videos and texts in the Web, guide books, observations.	Encouraging wild idea generation and the comparison of viewpoints. (P2) Requiring improvement in discourse quality because of pupils' unclear or irrelevant messages in the forums. (P5)
Cognitive	The modeling of inquiry strategies with a visual progressive inquiry cycle. Explication of the process progression through cumulative written guidelines in the virtual spaces.	Repeatedly connecting the phases of the inquiry cycle to concrete ongoing activities. Teaching pupils to read and follow written guidelines instead of leaning on the teacher. Stopping the process for self-evaluation and corrective actions when needed.

Pedagogical Practices in Case 2: School Children's Web Journal

As in Case 1, in Case 2 the teacher modeled the progression of the publishing process by dividing it into clear sub-tasks and drawing a picture of it on the blackboard. All the pupil groups did the tasks in the same order, but they worked at their own pace. During the same lesson, one group might still be writing their story while another group was already taking photos. Therefore, the teacher did not often give general instructions to the whole class; instead, he guided each group separately in the new procedures when they reached a new phase in their process. The following phases of the process were identified:

1. *Orientation to the task* by examining existing journals and forming editorial groups.
2. *Brainstorming of article topics* first in groups and then in a whole class discussion. After choosing a topic, the groups planned the content of their story by constructing a mind map in paper.
3. *Creating articles through process writing*, including multiple revisions of article drafts in groups and reciprocal feedback sessions between two groups.
4. *Taking photographs for the articles*, where the groups took photographs for their journal stories after the teacher had instructed them in the use of a digital camera. Each group also moved the photographs from the camera to a computer and further to the publishing tool with the teacher's support.
5. *Continuing the writing and publishing of articles*, where the groups finalized their stories, after which the teacher guided each group as they combined their story texts and photos into an article in the school's Web journal using the publishing program.
6. *Self-evaluation of the process and outcomes* included a session where pupils evaluated both the articles and the whole working process first in groups and then in a whole class discussion. Each pupil was also asked to evaluate his or her own contribution to the process.

In all, 261 separate sections representing a pedagogical design solution or scaffolding activity of Teacher 2 were chosen from the 12 lessons of Case 2. Because each pupil group worked at their own pace in the process, the chronological progression of the teachers' pedagogical activities cannot fully be presented according to the process phases. Most groups also worked on the writing project during their leisure time between the weekly lessons, which partly explains the variation in the process progression. The sections were distributed among the four support categories in the following way: 31% technical, 21% social, 25% epistemological and 23% cognitive. In Figure 2, the process is divided into weekly sessions consisting of two lessons, and the various process phases that separate pupil groups focused on during each session is mentioned in parentheses.

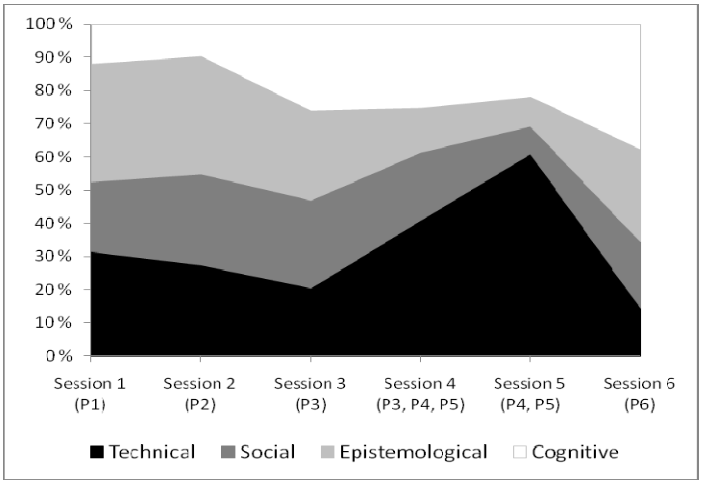


Figure 2. The Relative Distribution of the Pedagogical Infrastructure Elements in the Sessions of Case 2.

Figure 2 reveals that in the Web journal case, the support for learning technical competencies was substantial, especially in the middle phases when pupils learned to use digital cameras and to transfer material from the writing application and the cameras to the publishing program. Epistemological elements were emphasized at the beginning of the writing process when pupils brainstormed the content of their stories, whereas the teachers' structuring of cognitive or metacognitive elements increased towards the end of the process through promoting evaluative and self-reflective activities. Support for social activities and collaboration was less evident, but the teacher did made it clear in all situations whether pupils were supposed to work in groups or as a whole class, and which pupils or groups were supposed to work with each other. In Table 2 is a selection of exemplary pedagogical practices by Teacher 2, which can be regarded as representing sophisticated, expert-like practices that are good examples for other teachers. Separate examples are presented for overall design solutions and for scaffolding activities. When a practice relates to a specific phase of the process, the phase is mentioned in parentheses.

In the post-interviews after each session, the teacher evaluated that the goals of the unit were achieved well; in some phases pupils managed even better than the teacher expected, and the technology did not cause problems. Some pupils appeared to require more guidance in engaging in group work and in coordinating activities done at home and at school. The peer-commenting activity was tested by the teacher for the first time as part of the writing process, and he thought that its design and scaffolding required improvement. Overall, the teacher believed that after finishing the unit, the pupils were able to continue writing stories in the school's journal on their own.

Table 2: Examples of sophisticated pedagogical practices in Case 2.

Support	Overall design	Scaffolding
Technical	Introducing versatile tools according to the nature of the activity (Notepad, Magazine Factory, digital camera, IVB)	Guiding new technical skills for each group when they needed them in their own process, not in a general manner.
Social	Directing the students to make stories in pairs or groups. (P1) Directing the groups to evaluate each others' drafts through peer reviewing. (P3) Smoothly combining group and whole class discussions both in the brainstorming and in the evaluation phase.	Emphasizing to pupils in the final evaluation session that in group work, the whole group should take care of each member's turn and contribution in doing a collective activity. (P6)
Epistemo-logical	The whole process simulated the practices of a professional editorial staff. The pupils were allowed to choose the topic for their stories according to their own interests. (P2)	Explaining to pupils that they are documenting their school's preserved history with the stories. Repeatedly encouraging the pupils to be creative and to make the stories richer and more interesting.
Cognitive	Explicating the phases of a publishing activity through a visual process scheme. Organizing an evaluation session at the end in which the final stories, group work and individual contributions were evaluated by the pupils.	Regularly speaking with each group about what phases of the process they have finished and what phases are remaining, and asking the groups to report the status of their work.

Discussion and Conclusions

In the present study, the classroom practices (pedagogical design solutions or scaffolding activities) of two elementary school teachers who are experienced in the pedagogical use of ICT were analyzed in detail to exemplify sophisticated ways of using ICT in school education. The classroom practices were categorized according to their supporting role in the pedagogical infrastructure: technical, social, epistemological or cognitive (Lakkala, 2010).

In both of the units, one goal defined by the teachers was to develop pupils' *technical* competencies. In the units, the young pupils were taught quite challenging procedures in combining the usage of various technological applications, and the ICT usage was embedded in a meaningful context of a goal-oriented complex task. The teachers' approach complied with the opinion shared by many researchers (Pruulmann-Vengerfeldt, Kalmus & Runnel, 2008; Tierney, Bond & Bresler, 2006; Erstad, 2010) that digital competence is developed best in settings that include rich and integrated use of various technical tools as well as a wide range of activities that are based on complex, open-ended tasks such as product development, the solving of multidisciplinary problems, project work or collaboration.

Both teachers gave a large amount of responsibility to the pupils themselves in their group task, but also constantly monitored their progression, and when they good reasons, required corrective actions from the pupils if the criteria of the work were not met. It appears that very high-level and clear criteria both for *social* behavior (shared responsibility and negotiations, a common goal, proper commenting practices etc.) and for the *epistemological* quality of the work (a focus on concrete outcomes, the relevance of the produced questions and explanations, quality requirements for publishable texts etc.) were directing the teachers' design solutions and scaffolding activities throughout the process. In the classroom situations, these experienced teachers constantly and very conscientiously took care of those issues that they believed to be important in the designed activity.

Both of the teachers also promoted the pupils' self-regulative competences concerning the working practices through various sophisticated methods representing *cognitive* support. White and Fredriksen (2005) divided methods supporting metacognitive development into conceptual tools (e.g., guidelines, models, templates or software scaffolds) and metacognitive tasks (explicit tasks for planning, monitoring and reflecting one's own activity). In both units, the teachers provided the pupils with *conceptual tools* by modeling the working strategies and process progression through using visual modeling, writing systematic guidelines and phase descriptions, explicating the goals for the activity or emphasizing the criteria for high-level ways of working. *Metacognitive tasks* were explicitly included as essential working phases in both units, such as evaluating and improving one's own writings, commenting on others' productions, or having group-level and whole-class discussions in which the outcomes and working practices were reflected on and assessed together.

One important pedagogical characteristic in both of the units investigated was that the activity designed by the teachers was based on a holistic and goal-oriented working model that created a meaningful context for separate sub-activities and lessons. What is noteworthy is that the working models that the processes followed did not represent any pedagogical model that focuses on *learning* (such as the models of collaborative learning, problem-based learning or discovery learning). Rather, they were models simulating authentic *professional* practices in some field: in Case 1, scientific inquiry practices, and in Case 2, professional publication practices. For instance, Chinn and Malhotra (2002) criticized the practices of typically scientific reasoning being taught in schools through 'simple inquiry tasks' instead of 'authentic scientific inquiry tasks'. The teachers of the present study trusted in the pupils' ability to conduct a complex, challenging working process including various technical, social and epistemological elements, which is even more significant taking into account the nature of the pupil groups: very young pupils in Case 1, or a large proportion of pupils with special needs in Case 2.

The analysis revealed that the experienced teachers planned well-organized activities in which various elements of a pedagogical infrastructure were closely attended to and smoothly integrated. The study confirmed our previous result (Ilomäki et al., 2006) that technological skills as such are not sufficient for teachers to develop their pedagogical practices with ICT, but other competencies for planning, organizing, structuring and guiding pupils' learning activities with ICT are needed. The results can be used to complement current training materials presenting experienced teachers' practices with guidelines where essential aspects of the practices are unfolded and emphasized. A research aim for the future is to apply this approach to other situations and test the effectiveness of the materials in teacher training courses, collegial tutoring practices and self-study settings.

References

Bielaczyc, K. (2006). Designing social infrastructure: Critical issues in creating learning environments with technology. *Journal of the Learning Sciences 15*, 301–329.

Chi, M.T.H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *Journal of the Learning Sciences, 6*(3), 271–315.

Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic reasoning in schools: A theoretical framework for evaluating inquiry tasks. *Science Education, 86*, 175–218.

Erstad, O. (2010). Educating the digital generation. *Nordic Journal of Digital Literacy, 1*, 56–70.

FICTUP Pedagogical scenario (2010a). Exploring growth factors. Applying inquiry learning and virtual forums. Retrieved March 3, 2011, from http://www.fictup-project.eu/index.php/Exploring_growth_factors._Applying_inquiry_learning_and_virtual_forums
FICTUP Pedagogical scenario (2010b). School children's Web journal. Publishing with Magazine Factory tool. Retrieved March 3, 2011, from http://www.fictup-project.eu/index.php/School_children%E2%80%99s_Web_journal._Publishing_with_Magazine_Factory_tool
Cuban, L., Kirkpatrick, H., & Peck, C. (2001). High access and low use of technologies in high school classrooms: Explaining an apparent paradox. *American Educational Research Journal, 38*, 813–834.
Gibson, S., & Oberg, D. (2004). Visions and reality of Internet use in schools: Canadian perspectives. *British Journal of Educational Technology, 35*, 569–585.
Hogan, K., & Pressley, M. (1997). Scaffolding scientific competencies within classroom communities of inquiry. In K. Hogan & M. Pressley (Eds.), *Scaffolding student learning: Instructional approaches and issues* (pp. 74–107). Cambridge, MA: Brookline Books.
Hyo-Jeong S., & Kim, B. (2009). Learning about problem based learning: Student teachers integrating technology, pedagogy and content knowledge. *Australasian Journal of Educational Technology, 25*(1), 101–116.
Ilomäki, L., Lakkala, M., & Paavola, S. (2006). Case studies of learning objects used in school settings. *Learning, Media, and Technology, 31*(3), 249–267.
Jones C., Dirckinck-Holmfeld, L., & Lindström, B. (2006). A relational, indirect, meso-level approach to CSCL design in the next decade. *International Journal of Computer-Supported Collaborative Learning, 1*(1), 35–56.
Kozma, R., & Anderson, R. (2002). Qualitative case studies of innovative pedagogical practices using ICT. *Journal of Computer Assisted Learning, 18*(4), 387–394.
Lakkala, M. (2010). *How to design educational settings to promote collaborative inquiry: Pedagogical infrastructures for technology-enhanced progressive inquiry*. University of Helsinki, Institute of Behavioural Sciences, Studies in Psychology 66. Helsinki: Helsinki University Print.
Lakkala, M., Lallimo, J., & Hakkarainen, K. (2005). Teachers' pedagogical designs for technology-supported collective inquiry: A national case study. *Computers & Education, 45*(3), 337–356.
Lakkala, M., Muukkonen, H., Paavola, S., & Hakkarainen, K. (2008). Designing pedagogical infrastructures in university courses for technology-enhanced collaborative inquiry. *Research and Practice in Technology Enhanced Learning, 3*(1), 33–64.
OECD (2010). *Are the new Millennium learners making the grade? Technology use and educational performance in PISA 2006*. Paris: OECD Publishing.
Paavola, S., Lipponen, L., & Hakkarainen, K. (2002). Epistemological foundations for CSCL: A comparison of three models of innovative knowledge communities. In G. Stahl (Ed.), *Computer support for collaborative learning: Foundations for a CSCL community* (pp. 24–32). Hillsdale, NJ: Erlbaum.
Pedersen, S., Malmberg, P., Christensen, A. J., Pedersen, M., Nipper, S. Graem, C. D., et al. (Eds.). (2006). *E-learning Nordic 2006. Impact of ICT on education*. Copenhagen: Ramboll Management.
Putnam, R., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning. *Educational Researcher, 29*, 4–15.
Pruulmann-Vengerfeldt, P., Kalmus, V., & Runnel, P. (2008). Creating content or creating hype: Practices of online content creation and consumption in Estonia. *Cyberpsychology, 2*(1), 1–8.
Putnam, R., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning. *Educational Researcher, 29*, 4–15.
Smeets, E. (2005). Does ICT contribute to powerful learning environments in primary education? *Computers & Education, 44*, 343–355.
Tierney, R. J., Bond, E., & Bresler, J. (2006). Examining literate lives as students engage with multiple literacies. *Theory Into Practice, 45*(4), 359–367.
White, B., & Fredriksen, J. (2005). A theoretical framework and approach for fostering metacognitive development. *Educational Psychologist, 40*, 211–223.
Yin, R. K. (2003). *Case study research. Design and methods* (3rd ed.). Thousand Oaks, CA: Sage.

Acknowledgments

The research work was partially supported by the Lifelong Learning Programme of the European Union, contract 141929-2008-LLP-FR-COMENIUS-CMP (the FICTUP project). The authors thank the teachers and their pupils for participating in the study.

Classroom Orchestration: The Third Circle of Usability

Pierre Dillenbourg, Guillaume Zufferey, Hamed Alavi, Patrick Jermann, Son Do-Lenh, Quentin Bonnard, Sébastien Cuendet, Frédéric Kaplan, CRAFT, Ecole Polytechnique Fédérale de Lausanne (EPFL)
Email: Pierre.Dillenbourg@epfl.ch, Guillaume.Zufferey@epfl.ch, Hamed.Alavi@epfl.ch, Patrick.Jermann@epfl.ch, Son.Dolenh@epfl.ch, Quentin.Bonnard@epfl.ch, Sebastien.Cuendet@epfl.ch, Frederic.Kaplan@epfl.ch

Abstract: We analyze classroom orchestration as a question of usability in which the classroom is the user. Our experiments revealed design features that reduce the global orchestration load. According to our studies in vocational schools, paper-based interfaces have the potential of making educational workflows tangible, i.e. both visible and manipulable. Our studies in university classes converge on minimalism: they reveal the effectiveness o tools that make visible what is invisible but do not analyze, predict or decide for teachers. These studies revealed a third circle of usability. The first circle concerns individual usability (HCI). The second circle is about design for teams (CSCL/CSCW). The third circle raises design choices that impart visibility, reification and minimalism on classroom orchestration. The fact that a CSCL environment allows or not students to look at what the next team is doing (e.g. tabletops versus desktops) illustrates the third circle issues that are important for orchestration.

Introduction

When CSCL was implemented on desktops, the physical organization of classrooms was rarely addressed as a research topic. The use of small mobile devices as well as large immobile devices (e.g. tabletops) brought forward the physical orchestration of the classroom. An early example was given by Roschelle and Pea (2002): when a student walks across the classroom to share PDA data by infrared instead of sending them wirelessly, this publicly visible walk provides the teacher and other students with the awareness of the actual dataflow. Making the educational workflow visible and tangible has an effect on classroom orchestration. A recent example comes from Nussbaum et al (2010). All kids in a classroom interact with a mouse on a single display. Each student owns a tiny subset of the display area, as small as a phone display. The very same learning activity could be conducted on a central display or on personal displays yet leads to completely different forms of orchestration. Analyzing the impact of CSCL design on classroom orchestration is the goal of this conceptual paper. Therefore, we first present a study where we used an augmented reality simulation in classrooms for two years. This simulation combines a tangible interface and a paper-based interface. The added value of tangibles was obvious in terms of individual usability and teamwork, but we also understood that the added value of paper concerned a third circle of usability: its integration in the classroom ecosystem. While paper-based computing is a good example of orchestration, we aim to go higher in abstraction. Therefore, in the second section, we define this third circle with concepts of orchestration, awareness and workflow. They are illustrated with new designs and new results in the last sections with paper computing as well as with other technologies.

Why is Paper Highly Usable in Classrooms?

The spread of reading devices led scholars to compare digital and paper documents in terms of readability or annotations. We analyzed paper interfaces from another perspective: classroom orchestration. Tinker is an augmented-reality simulation for training apprentices in logistics (Zufferey et al, 2009). Teams build the mock-up of a warehouse by placing shelves on a table (Fig. 1 left). The TinkerLamp includes a camera and a beamer. It recognizes the visual markers on the shelves, computes a model of the warehouse and displays information on the table and on the shelves (their contents, the movement of the forklifts, the surfaces used, etc.).

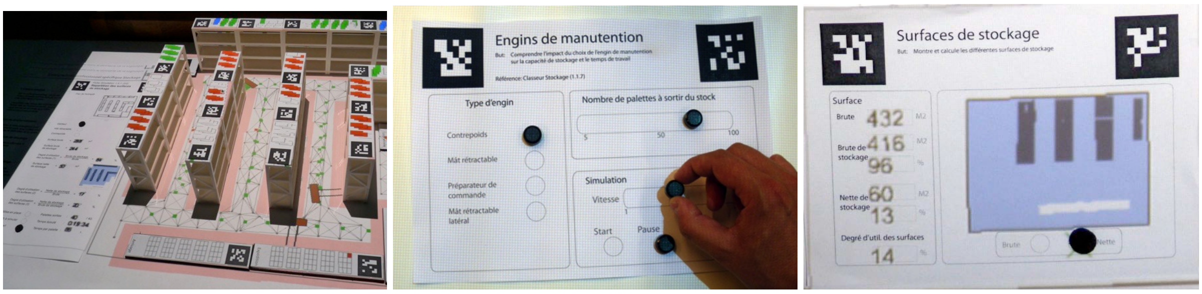


Figure 1. Input (Middle) and Output (Right) Sheets for a Tangible Simulation (Left).

Empirical studies revealed that this tangible interface outperformed a multi-touch table running the same simulation (Lucchi et al, 2010). Tinker has been used in 10 classes from 5 schools, through several studies, but in the beginning, teachers did not feel comfortable in setting up the activities and controlling the simulation. After several prototypes, we found that interactive paper sheets facilitated their interactions. Teachers and apprentices interact with the system by placing tokens on the sheets (Fig.1 middle and right). The camera identifies the sheet and its orientation thanks to the markers, retrieves the coordinates of the input area and identifies the black tokens as equivalent to mouse clicks. On these sheets, the system beams information such as performance measures (e.g. the average time to bring a box from the shelves to the truck) and a reduced map of the warehouse. These paper sheets are the equivalent of menus and palettes in WIMP interfaces but are persistent, i.e. they remain visible outside the interaction area, which has a huge impact on usability for individuals and teams. We do not analyse these individual usability advantages here but consider how paper influences orchestration, i.e. from the teacher perspective.

The tangible interface enabled students to explore rapidly many warehouse designs, but tinkering is not learning. Learning required reflective activities that have to be enforced by the teacher. One student in each team (usually four teams per classroom) had to copy the warehouse layout and performance values by passing a pen on the beamed information, to bring this sheet to the blackboard and to copy the information again on the board (Fig.2 left). The teacher then asks the students to compare their results and to explain why a particular design was better than another (Fig. 2 right). While a client-server architecture would display the same data faster, this media discontinuity affords richer forms of classroom organisation: *handling paper makes the workflow visible and tangible.*

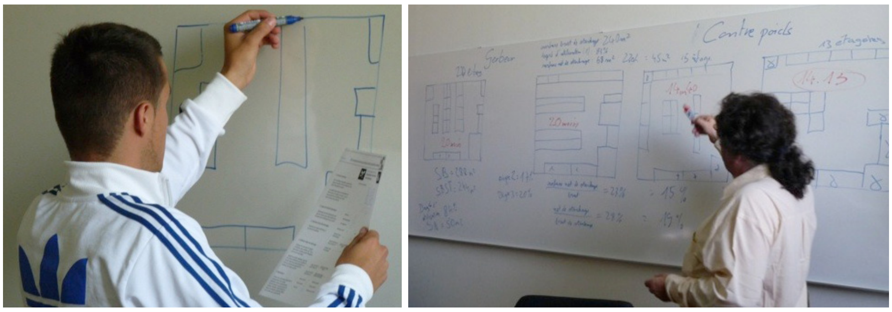


Figure 2. Each team reports their results by using annotated sheets. Comparing the four solutions, the teacher pushes students to discover the properties of an efficient warehouse.

After a few years of collaboration with teachers, the curriculum has been re-structured on Tinker sheets (Fig. 3). Concretely, this curriculum has the form of an A4 binder. To run an activity defined in the curriculum, the teacher selects a sheet and places it partly under the lamp. The part that is viewed by the camera contains all the information that the system needs in order to run the learning scenario associated to that page. After the lesson, the teacher may annotate this sheet with comments for the next year, make copies, etc. Paper makes tangible the educational design cycle: prepare the lesson – teach – reflect.

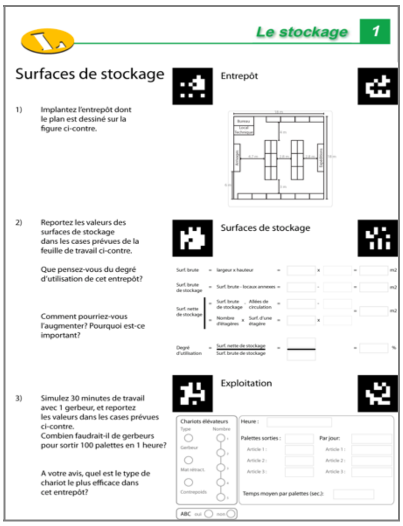


Figure 3. Curriculum Sheets: the Right Part (with Visual Markers) is Input/Output.

These examples revealed an interesting feature of paper-based computing: paper sheets *make the educational workflow tangible*, which implies that this workflow becomes visible and modifiable. Distributing sheets, collecting them, storing them or annotating them are common practices in school ecosystems. Contrarily to the myth of the paperless office, we argue that paper-based interfaces are well suited for *routine worlds*, i.e. environments where the main activities are recurrent and where users develop solid habits.

Circles of Usability

The relevance of paper-based computing for orchestration has only been presented here as an example. We aim to abstract and generalize these findings to a broader range of CSCL technology. We therefore define *three circles of usability*. The first circle concerns the understanding of how individuals interact with (learning) environments. For instance, the tangible shelves of the TinkerLamp enable faster manipulations than a multi-touch table, they provide a real 3D perspective and they off-load students from tackling the scale issues they face when drawing warehouses. The second circle concerns team processes: how do collaboration tools shape interactions among learners; CSCL became paraphrased as *design for conversation* (Roschelle's,1992). The third circle concerns the integration of CSCL environments in the classroom or *design for orchestration*.

What is considered as '*the user'* is an individual person at Circle 1, a team at Circle 2 and the classroom at Circle 3. Referring to *the classroom as the user* means that we aim to understand its processes and constraints. At Circle 1, the constraints were the individual's cognitive load, background knowledge, experience, motivation, etc. At Circle 2, the constraints were related to the team's need to build enough shared understanding to carry the task at hand, the peers' level of interdependence, etc. At Circle 3, teachers have to cope with many constraints: curriculum relevance, time budget, time segmentation, physical space, discipline, security, etc. Understanding the relationship between CSCL design and the management of these constraints is what we refer to as *usability at the classroom level*.

To understand the design features that make a tool 'usable' at the classroom level, we use three concepts: orchestration, awareness and workflows. Before describing them, we need to clarify two points. First, there also exists several circles around the classroom (the school, the community, the society, etc.) that we do no describe here: they receive a lot of attention in CSCL while the classroom circle has been neglected. Second, the term "classroom" is used as a flag: it does not exclude activities outside the classroom (field trips, museum visits, homework, etc.) or corporate training (workshops, seminars, etc.). However, our analysis is restricted to situations where a person (teacher, teaching assistant, parent, workplace supervisor, etc.) has the *responsibility* to bring other persons to reach learning goals (Hoppe, personal communication). For the sake of simplicity, we refer to this responsible person as the teacher. This paper is hence not addressing informal learning.

Orchestration

What does a teacher do if he conducts a CSCL script designed for teams of three, with three roles per team, when suddenly one student drops out the class? What if the next script activity is 40 minutes long but the class ends in 30 minutes? What if two students refuse to work together? Classroom *orchestration* refers to the real time management by a teacher of multiple learning activities within a multi-constrained environment. Classroom management is as old as schools, but it became salient in CSCL when scenarios (or scripts) began integrating individual (e.g. reading), collaborative and class activities (e.g. readings, lectures). This integration requires adapting the script on the fly in order to cope with many constraints. We enumerated many of them (Dillenbourg & Jermann, 2010):

- *Curriculum constraints*: how relevant is the topic with respect to the learning objectives listed in the curriculum? Do students have the prerequisites? Etc.
- *Assessment constraints*: are my learning activities compatible with exams? Does my CSCL tool require a reasonable workload? Etc.
- *Time constraints*: how much time is necessary? How much time is left before the break and how much flexibility do we have around these two factors? How much time is lost simply to install the tool? Etc.
- *Sustainability constraints*: how much time and energy must teachers engage to prepare and run this method? How long can they do it? How much does it cost? Etc.
- *Space constraints*: do I have the space necessary in my classroom do set up these activities? Can I move furniture? Can I walk around the classroom? Is there enough daylight? Etc.
- *Discipline constraints*: Can I keep control of my class? Is the level of noise in the classroom below what is tolerated by the school director? Etc.

These constraints could be considered as practical problems, poorly related to learning theories. It is true that they typically correspond to factors that we treated as "controlled variables" when conducting field studies. However, ignoring these factors probably explains why CSCL has difficulties in scaling up from field experiments to broader educational impact: they not explain why students learn, they may explain why a method could fail. This paper illustrates how CSCL could pay more attention to design features that allow teachers to manage multiple classroom constraints.

Awareness Tools: Less Ambitious than Student Modeling

Orchestration could be described as a regulation loop: the teacher monitors the classroom, compares its state to some desirable state in the scenario, and adapts the scenario accordingly. This loop defines two points of orchestration: *state awareness* and *workflow manipulation*.

Individualizing a pedagogical scenario relies on student modeling: the system (or the teacher) aims at inferring from the student's behavior what (s)he has understood and not understood. This in-depth understanding of students is not possible when orchestrating classroom activities with 25 students. The analysis of traces in CSCL tools has progressively lost depth and gained breadth, for instance proposing visualizations of conversation patterns without analyzing the semantics of utterances (Bachour et al, 2010). This evolution brings us closer to the CSCW notion of "awareness tools" (Greenberg, Gutwin & Cockburn,1996): informing users about the activity of their co-workers: who is on-line, on which document or paragraph are my peers working on, are they available, etc.? Awareness is less ambitious than student modeling since it shares behavioral information among users without cognitive diagnosis. The information overload that awareness may trigger has been tackled through the concept of 'filtering', i.e. selecting the relevant information to share. By downgrading 'student modeling' to 'awareness', we stress the need for *minimalism* in the design of orchestration technologies.

Educational Workflows: The Light and Dark Sides of Integration

Let us illustrate workflows in the context of *"integrated learning*" scenarios, i.e. CSCL scripts that combine team learning with individual activities and class-wide activities. One example is ArgueGraph, which scaffolds argumentation by forming pairs of students with conflicting opinions. This script includes the following phases: (1) students individually answer a questionnaire; (2) the system plots them on a 2D opinion map based on their answers (the teacher having previously defined an X and a Y value for each answer); (3) the system forms pairs of students based on their distance on the graph; (4) pairs of students answer the same questionnaire again; (5) the teachers conducts a lecture based on all answers collected by the system during phases 1 and 4. The individual, group and class-wide activities are computationally "integrated" because the data produced in an activity are necessary inputs for another activity. For instance, in ArgueGraph, Phase 1 answers are used by the system to build the map in Phase 2 and to make teams in Phase 3 as well as for the debriefing (Phase 5). The notion of workflow – another CSCW concept – fits well with CSCL scripts because it does not only refer to a sequence of activities, but also to the underlying flow of data across these activities.

Since they run in the 'back office' of scripts, workflows could be seen as a technical rather than a pedagogical issue. This is not the case: workflows constitute both the strengths and weaknesses of CSCL scripts. Regarding the strengths, a workflow is the condition for integrating heterogeneous activities in a consistent whole. As for weaknesses, workflows usually are internal to the software (not accessible from outside except as log files) and 'hard wired', i.e. difficult to modify. In other words, workflows both enable the execution of CSCL scripts and reduce their flexibility, which creates new constraints to teachers.

A few years ago, we conducted the ArgueGraph script with paper instead of computers and it worked surprisingly well. We distributed the questionnaires as paper sheets. When all students had completed it, we gave them the scheme for scoring their answers in the same way the ArgueGraph software did. They communicated their graph position to the teacher who plotted them on the blackboard and formed pairs. The extra work for manually counting answers and forming pairs was compensated by the ease of manipulation. Nonetheless, we lost a key functionality, the collection of students' justifications for the final debriefing. To avoid this, recent developments in 'paper computing' (next section) combine the advantages of an executable script with the advantages of manipulating real paper.

Paper Computing: Making Educational Workflows Tangible

Tinker illustrates that paper implements tangible and visible workflows: as explained in the first section, paper sheets pass from one activity to the next one; they get annotated, distributed, shared, etc. An additional example concerns the classroom-homework-classroom workflow. In Figure 4, the apprentices save the warehouse layouts they have designed and select what they consider as their four best layouts. The system generates an "individual fieldwork sheet" that the apprentices print and take away: they have to compare these saved layouts to the warehouse in which they work in order to connect school knowledge with experience. Since this printed sheet has the same tags as the other TinkerSheets, they can be used as input for the next school activity.

A second orchestration example addresses a well-known problem in learning from simulations: students can run a simulation many times without much reflection (De Jong & van Jooligen, 1998). As a tangible interface made our simulation especially playful, this risk is high in Tinker. We therefore developed the *paper orchestration keys (POKs).* The "simulate" POK is used to force hypotheses*:* teams cannot run the simulation without showing this POK to the camera. The standard scenario is that the teacher has the key in hands when walking from one team to another and that apprentices hence have to call him when they want to run the simulation. Before giving the key, the teacher will for instance ask them to predict if the warehouse performance (average time to move a box from the shelves to the truck) will be higher or lower than in the previous run, and to explain why. The key empowers the teacher in his management of teams and makes the scenario easy to modify: the teacher may decide to leave a copy of the key to a good team, to give a key to all teams, to take back a key, etc. This could be achieved with options in the software interface but the paper key makes these workflow changes visible for all actors. The teacher and all other teams see at any time who has the key and who does not; they can take it or give it much faster than by tuning options in sub-menus of an application. Note that this POK empower teachers in orchestrating constructivist activities, not in lecturing.

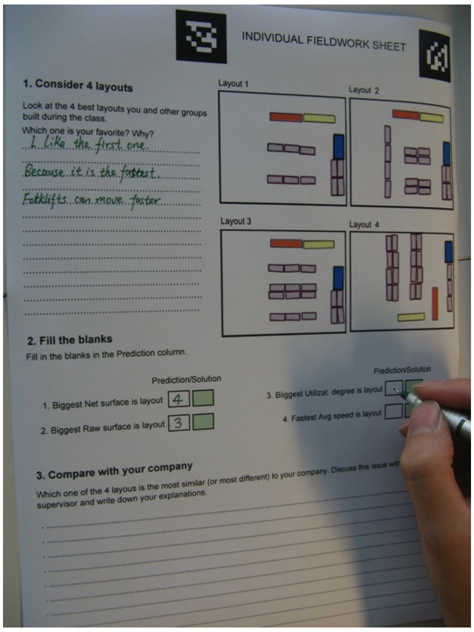


Figure 4. The Classroom Activity Generated Homework Sheets That Will Be Reused by the Environment.

POK s have also implemented in an augmented reality CSCL environment (Figure 5, left) that uses paper to teach geometry in elementary schools (properties of triangles/quadrilaterals, surfaces, angles, symmetry axes,…). The research question is concerned with the effectiveness of learning activities that use paper sheets as tangible objects: paper-made polygons can be rotated, folded (axes), cut, etc. Other types of paper sheets are used as operators. For instance, the students could show a card to overlay a grid over a paper triangle in order to estimate its surface by counting squares. While geometrical objects are made of simple paper, operations are printed on POKs similar to collectable playing cards. Teachers use these POKs to orchestrate the activity in different ways: they may show a POK to the system to display the length of each segment (providing feedback to the kids); they may decide to provide students with quantitative POKs (e.g. measuring the surface) only after they qualitatively understood the notion; they may distribute different POKs to different members of a team to define roles, etc. POKs have to be more robust than sheets representing polygons because they are used for longer periods (paper polygons are not reused since they are cut, folded, colored, etc. by the students) and more rigid to be as easily manipulated as play cards.

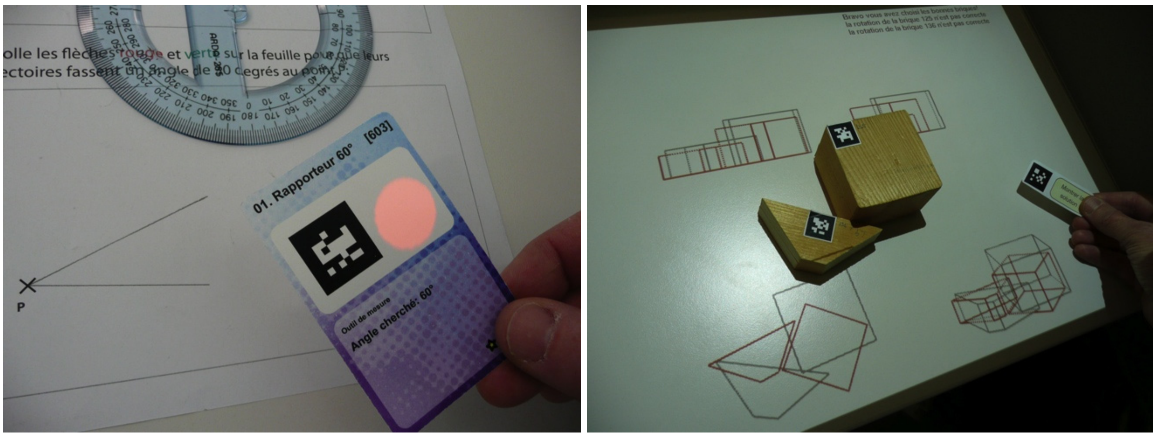


Figure 5. Left: this card asks for feedback on the angle to be constructed (the beam red dot indicates it is not correct). Right: this card provides scaffolds (as explained below).

The same approach is used in another environment: for training apprentices in carpentry (Figure 5, right). The research question here relates to the development of spatial reasoning skills with augmented drawings. The apprentices manipulate wooden blocks and the computer displays their projections in the three orthogonal planes (what they have to draw at school). An example of a script is that a team "saves" a construction and gives it to another team that has to assemble the blocks in a way that matched first team's projections. The POKs presented by the teacher displays (in red) a scaffold for the second team: the difference between their current construction and the one they have to produce.

These examples illustrate that paper interfaces make the workflow tangible and a tangible workflow is visible to all actors and easier to modify. Paper-based interfaces are promising tools to combine this tangibility / visibility without abandoning computational power. Of course, we do not pretend that paper intrinsically facilitates orchestration; it is a matter of design. For designing interfaces, we propose a simple model (PAW). Interactive paper sheets are covered by three layers of information: those printed in advance (P), those beamed by the augmented reality environment (A) and those written or drawn by the learners and/or the teacher (W). Designing paper-based interfaces is about understanding the complementarily of the 3 layers. The P layer contains hard elements from the script, the A layer makes the script interactive but A info is lost when kids leave the system while the W layer will remain after the session The need to produce tangible traces of learning activities (e.g. for parents) typically is a classroom constraint that CSCL did not pay much attention to.

Modest Computing: Minimalism in Ambient Awareness

So far, we stressed the usability of paper for orchestrating activities. To broaden our argument, we now present a very different orchestration tool. It originates from our observations of teamwork at the university level. Typically a first year course in physics is composed of two hours of lecture plus two hours of exercises per week. During these exercises, most students work in small groups (two to four students) on a list of 8-10 exercises. When students are stuck, they raise their hands and one of the teaching assistants (TAs) comes when (s)he is available. In terms of orchestration, this is fairly simple compared to complex CSCL scripts, yet it is far from being optimal. Twelve recitations sections have been videotaped from three different courses involving around eighty students altogether (Alavi et al, 2009). While waiting for the TA, students spend 62% of their time visually chasing the TA because, if they do not grab him or her as soon as (s)he is available, (s)he might go to another team. Other problems were observed such as unanswered questions (students give up) or the TA helping a team that has been waiting much less than another one.

Alavi designed two tools to address these problems. The first one, named Lantern (Fig. 6 left) is a small device (in size of 0.5 L bottle) consisting of five LEDs installed on a stub-shape PCB and covered by a blurry plastic cylinder with one microprocessor to control the LEDs. By turning the cover, students indicate which exercise they are working on: every colour corresponds to one exercise. The height of the colour bar indicates how much time that has been spent on the current exercise. When a team wants to call the TA, it presses the Lantern which starts blinking. The blinking rate increases slowly indicating the waiting time. The second tool, named Shelf (Fig. 6 right) uses exactly the same visual codes as Lantern, but students communicate with a clicker and the status of teams is displayed centrally on a display. We provided both tools to two courses of physics. In both classes, students and TAs used Shelf for three weeks, after that they switched to use Lantern for four weeks. In total, Shelf has been used for around 12 hours and Lantern for 14 hours. The main result is that the estimated time wasted in chasing the TA was reduced from 62% in our early observations to 16% in the Shelve condition and to 6% in the Lantern condition. Students simply continue to work while waiting.

Figure 6. The Lantern Device (Left) and the Shelve Environment (Right).

When orchestrating 10 teams of students, the TA iteratively faces two questions: which team should I help now and what should I tell them. Shelf and Lantern are only concerned with the first question, which is easier than the second one. Moreover, these tools do not decide where the TA should go next. They more modestly provide TAs with some "awareness," as defined earlier, of the teams' behaviour. They are not smart tools; they neither interpret activities nor predict the need to intervene, but they *simply make visible things that would otherwise remain invisible*: working time and waiting time. The decision remains in the TA's hands. Our *minimalism* does not only apply to the functionality of Lantern but also to its design. *We deliberately reduced the resolution of the display:* instead of displaying the precise exercise number and the exact waiting time, Lantern provides degraded information. The term "ambient" is used for displays that do not monopolize the visual attention of users. Given the main effects of the Lantern, we are even tempted to believe that this minimalism is a condition for orchestration, but this is only a hypothesis at this point.

The second interesting result in terms of orchestration is that the physical layout had an impact on the social processes. Shelf induced some competition between teams, while Lantern triggered collaboration between teams: when Team could see that a neighbouring Team was moving to a next exercise (they changed colour), Team 1 would sometimes ask Team 2 for a suggestion. Lantern generated a social/spatial organisation of the classroom into spatial clusters of two to three teams. The fact that t two tools that provide almost the same information generate different social processes illustrates the physicality of orchestration: it is about mobility, gaze directions and distances between all classroom actors.

Let us analyze more deeply the fact that teams peripherally perceive their friends, hence see when a neighboring team moves to the next exercise, which triggers inter-team interactions. Similarly, in the Tinker studies, students did also look over the shoulders of other students to see the warehouse being built by other teams. Unlike desktops, tabletop environments induce indeed two interaction spheres: a first sphere of users who manipulate objects on the table and a second sphere of students who can see or who can hear what is done in the first sphere. "*Looking over the shoulder*" had a positive effect in the Lantern study but could have a negative effect in the Tinker classrooms (students copying the warehouse of others instead of exploring). Whether they are deliberate or accidental, "looking over the shoulders" and "over-hearing" often happen in classrooms. They are realities of orchestration, not investigated in CSCL, that illustrate well the third circle of usability. Let us review the three circles in terms of what users visually perceived. At Circle 1 of usability, HCI is concerned by how well the user perceives the display (readability, understanding of symbols, etc.). At Circle 2, CSCL/CSCW investigated if team members should or not perceive the same things (WYSIWIS: "what you see is what I see"). At Circle 3, a new concern is to analyze when team members look at the display of another team. The same circles can be defined for audio perception, "overhearing" being at Circle 3.

Figure 7. Three Designs of the TinkerLamp: the Used Model (Left) and Two New Models.

Investigating "orchestration" requires an analysis of how CSCL designs influence "looking over the shoulder" and "overhearing", namely how they modify the line of sight for students and, very importantly, for the teacher. Let us illustrate this with the design of the Tinker Lamp. In the reported experiments, the teacher placed four lamps (Fig. 7 left) in the classroom. There were several problems in putting the beamer above the surface. Therefore, our new designs include the beamer on the table and a mirror above the table (we do not project from below the table because we have to display information on the shelves and on paper sheets). The two new lamps are illustrated in Fig. 7 middle and left. Their designs induce different orchestration processes. The black model prevents the teacher from seeing in a glance what students are doing while the white model does not break the line of sight. The white model is better suited for the scenario we used in logistics training where teams simultaneously use up to five Tinker in the classroom. Orchestrating a classroom with five black lamps would require the teacher to run around the class for monitoring what students do. Conversely, elementary classrooms have a corner with a bookshelf, a sofa and a table with a computer. The black model is better suited these classrooms: two students can work there without perturbing too much the rest of the class.

Conclusions

For several years, CSCL has been striving for a better integration of our tools into educational ecosystems: CSCL does not appear in a vacuum but is part of an ecosystem and should be integrated into other activities (individual and class-wide), with or without computers, inside or outside the classroom, etc. This evolution reflects the maturity of CSCL (team learning is not the unique approach) as well as the technological evolution (mobile devices, tabletops, etc.). This integration requires deepening our understanding of orchestration.

This paper addressed the question "how do CSCL tools influence orchestration?" from a perspective that may be shocking: we stressed the practical aspects of activities rather than the learning processes themselves. But our experience is that the success of CSCL tools is hidden in these implementation details. Ignoring them systematically leads to develop environments that increase the teacher's "global orchestration load." Every menu to pull down and option to select increases the workload of a teacher who acts live in front of 20 or 30 students. The experiments we conducted led to *simple design principles*: (1) strive for *minimalism* in design (few functionalities, reduced resolution of information), (2) *care for visibility* by taking care of lines of sight and implement *reification* (make visible aspects that are usually invisible) (3) make the educational *workflow tangible* (4) *empower teachers.* The last point is the consequence of the previous ones. Empowering teachers is neither wishful thinking, nor pushing authoritative visions of education. What we mean is that, in CSCL environments, the teacher should literally hold the scenario in his or her hands, such that it is easy to manipulate, and not simply have a few options to select in a predetermined script before unfolding him or her. The evolution of computer science provides us with new tools to embed these principles in CSCL environments: paper-based computing, tangibles, ambient displays, tabletops.

These "*design for orchestration*" principles do not form a theory at this stage. Nonetheless, as one step in that direction, we proposed the *third circle of usability*, i.e. a set of concepts that are part of orchestration and that have not been considered so far at lower circles of orchestration. By using somehow provocatively the word 'usability', we do not discard that classroom orchestration raises other pedagogical factors, but we stress that fact that this classroom usability is a necessary condition to implement effective learning scenarios.

References

Alavi, H. S., Kaplan , F. & Dillenbourg, P. (2009) Distributed Awareness for Class Orchestration. In U. Cress, V. Dimitrova & M. Specht (Eds). Proceedings of the 4th Europen Conference on Technology Enhanced Learning, pp. 211-225, EC-TEL 2009, Nice, France, Sept. 29th – Oct. 2nd. Springer, LNCS 5794

Bachour, K., Kaplan, F., Dillenbourg, P. (2010) "An Interactive Table for Supporting Participation Balance in Face-to-Face Collaborative Learning," *IEEE Transactions on Learning Technologies*, 3 (3). 203-213

De Jong and van Jooligen (1998). Scientific Discovery Learning with Computer Simulations of Conceptual Domains. *Review of educational research*, 68, pp. 179-202.

Dillenbourg, P. & Fischer, F. (2007). Basics of Computer-Supported Collaborative Learning. *Zeitschrift für Berufs- und Wirtschaftspädagogik.* 21, pp. 111-130.

Dillenbourg, P. & Jermann, P. (2010) Technology for Classroom Orchestration, in: M. S. Khine & I. M. Saleh (Eds) *New Science of Learning: Cognition, Computers and Collaboration in Education* (pp. 525 – 552). Springer. Dordrecht

Greenberg, S., Gutwin, C. & Cockburn, A. (1996) Awareness Through Fisheye Views in Relaxed-WYSIWIS Groupware. Proc. of *Graphics Interface* (pp. 28-38), Toronto, CA, May 21-24. Morgan Kauffman.

Lucchi, A., Jermann, P., Zufferey, G., and Dillenbourg, P. (2010). An empirical evaluation of touch and tangible interfaces for tabletop displays. In *Proc.s of the 4th int. conf. on Tangible, Embedded, and Embodied interaction* (Cambridge, MA, USA, January 24 - 27, 2010). TEI '10. ACM, New York, NY, 177-184.

Nussbaum, M. C. Alcoholado, A. Tagle, F. Gomez, F. Denardin, H. Susaeta, M. Villalta, K. Toyama (2010), One Mouse per Child: Interpersonal Computer for Personal Formative Assessment, Submitted

Roschelle, J. & Pea, R. (2002) A walk on the WILD side: How wireless handhelds may change computer-supported collaborative learning. *International Journal of Cognition and Technology*, 1(2) pp.145-168.

Roschelle, J. (1992) Learning by Collaborating: Convergent Conceptual Change. *J. of the Learning Sciences*, 2, 235-276.

Zufferey, G., Jermann, P., Lucchi, A., and Dillenbourg, P. (2009) TinkerSheets: using paper forms to control and visualize tangible simulations. In *Proceedings of the 3rd international Conference on Tangible and Embedded interaction*(Cambridge, UK, Feb, 16 - 18, 2009). TEI '09. ACM, New York, NY, 377-384.

Acknowledgments

Tinker is developed within the leading house "Technologies for vocational education" supported by the Swiss Ministry of Economical Affairs. Lantern has been funded by the Swiss NSF under the ProDoc PDFM1-118708. Thanks to O. Guédat and W. Hokenmaier who built the hardware as well as to K. Bachour, J. Kurzo, A Ryser, P. Ogay and the teachers who co-designed these tools. Thanks to C. Sanchez, M. Chablais, S. Testuz,

Mind the Gaps: Using Patterns to Change Everyday Classroom Practice Towards Contingent CSCL Teaching

Luis P. Prieto, Sara Villagrá-Sobrino, Yannis Dimitriadis, University of Valladolid, SPAIN,
Email: lprisan@gsic.uva.es, sarena@pdg.uva.es, yannis@tel.uva.es
Patricia Schank, William Penuel, Angela Haydel DeBarger, SRI International, Menlo Park CA, USA,
Email: patricia.schank@sri.com, william.penuel@sri.com, angela.haydel@sri.com

Abstract: Educational research has established the benefits of adapting lesson plans and teaching to the evolution of student knowledge and emergent occurrences in the classroom. This kind of improvisatory adaptation of teaching, however, is seldom seen in everyday classroom practice of CSCL. This paper describes two independent research projects that aim to influence authentic classroom practices to promote this kind of teaching using the same collaborative learning tool, Group Scribbles. Evidence from both projects shows that merely providing technologies that support this adaptation is not enough to provoke the change and that exposing teachers to good uses of the tool (in the form of pedagogical patterns) also has limited success. Both projects highlight the difficulty for practitioners to bridge the gap between de-contextualized advice and contextualized classroom situations. We propose the use of more atomic, actionable moves to help teachers orchestrate technology to support deeper collaborative knowledge building.

Introduction

A main concern of the CSCL community in recent years (as it is evident from Dillenbourg, 2009, and from the title of this CSCL 2011 conference) is how to apply CSCL research results to everyday educational practice. For example, facilitating highly interactive discussion and adjustment of instruction based on what teachers learn from student responses have long been considered beneficial (Alexander, 2008). This is especially true for collaborative learning, where the teacher's role shifts to facilitator, working with students to help them tackle the learning challenge through the sharing and construction of questions, ideas, or data. However, the adjustment of teaching for deep and meaningful collaborative discussion is seldom seen in everyday teacher practice with information and communication technologies (ICT) (Kennewell et al., 2008).

The success of such agile instruction depends on teachers having resources that help them adjust lessons, contingent on what they find out students know and can do (Beatty et al., 2006). Yet existing research is insufficient to indicate how to best scaffold improvisational adjustments in everyday classroom practice, especially in reaction to rich, constructed student responses.

Pedagogical patterns are a common resource for scaffolding teacher enactment and enabling contingent teaching with classroom network technologies (DeBarger et al., 2010; Conole et al., in press; DiGiano et al., 2003; Prieto et al, 2010). These patterns represent best practices and prior knowledge by expert practitioners as tried-and-true solutions to recurrent problems in a field or practice (Alexander et al., 1977). Patterns can be formulated to help teachers make student thinking visible and engage students in genuine dialogue so that the students shape the flow and direction of the discussion. Such dialogic discussions improve students' learning and development of scientific explanations (Nystrand & Gamoran, 1991).

Technology is another resource for supporting instruction that can flexibly adjust in response to student thinking, making it easy for the teacher to modulate the discussion by, for example, posing a new question or activity on the fly based on student responses (Penuel et al., 2005, Roschelle et al, 2007). Only recently have the appropriate kinds of flexible technologies emerged to support such malleable, interactive instruction. In the context of collaborative learning, clickers have foreshadowed this new potential, and virtual whiteboard systems with individual and shared displays show promise in supporting the rapid exchange and submission of sketched representations (Anderson et al., 2007). Group Scribbles, a collaborative tool based on the familiar metaphors of private/public boards and adhesive stickers, is an example of a flexible, collaborative whiteboard system that supports a range of learning activities (DeBarger et al., 2010; Dimitriadis et al., 2007; Looi et al., 2010).

This paper relates findings from two independent projects—one in Spain and one in the United States––that studied Group Scribbles implementations in real classroom settings to better understand the impact of flexible network technology and pedagogical patterns on contingent, student-centered teaching. Both projects employed professional development workshops and classroom observations, but otherwise used different research approaches. Using a largely bottom-up approach, researchers in Spain introduced Group Scribbles into primary school classrooms and helped teachers transform their lesson ideas into Group Scribbles activities, documenting such practices as common design and enactment patterns and improvisational adjustments to instruction. Using a mix of researcher- and co-designed pedagogical patterns and interactive assessment activities, researchers in the United States introduced Group Scribbles to middle school teachers, documenting implementation challenges and how teachers adjusted instruction within the provided structure.

Evidence from both projects shows that a key challenge in putting contingent teaching into practice, and especially in reusing best practices and prior knowledge from research (e.g., in the form of pedagogical patterns), is bridging the gap between de-contextualized advice and specific classroom situations and actions (Goodwin & Duranti, 1992). The multiple decisions to be made when implementing a pedagogical pattern or when reacting to rich student-generated responses present pedagogical pitfalls and risks. To circumvent this problem, researchers on both projects independently proposed using more atomic and actionable instructional moves. Initial evidence from both projects suggests that by using this approach (rather than just de-contextualized pedagogical patterns), the design and enactment of activities is greatly facilitated.

This paper first describes the research approaches of both projects and the main findings regarding the implementation of contingent teaching. Then the challenge of bridging the gap between de-contextualized advice and contextualized enactments in everyday practice is discussed. Finally, conclusions about implementing innovations for contingent teaching in CSCL are presented, as are paths of potential future work.

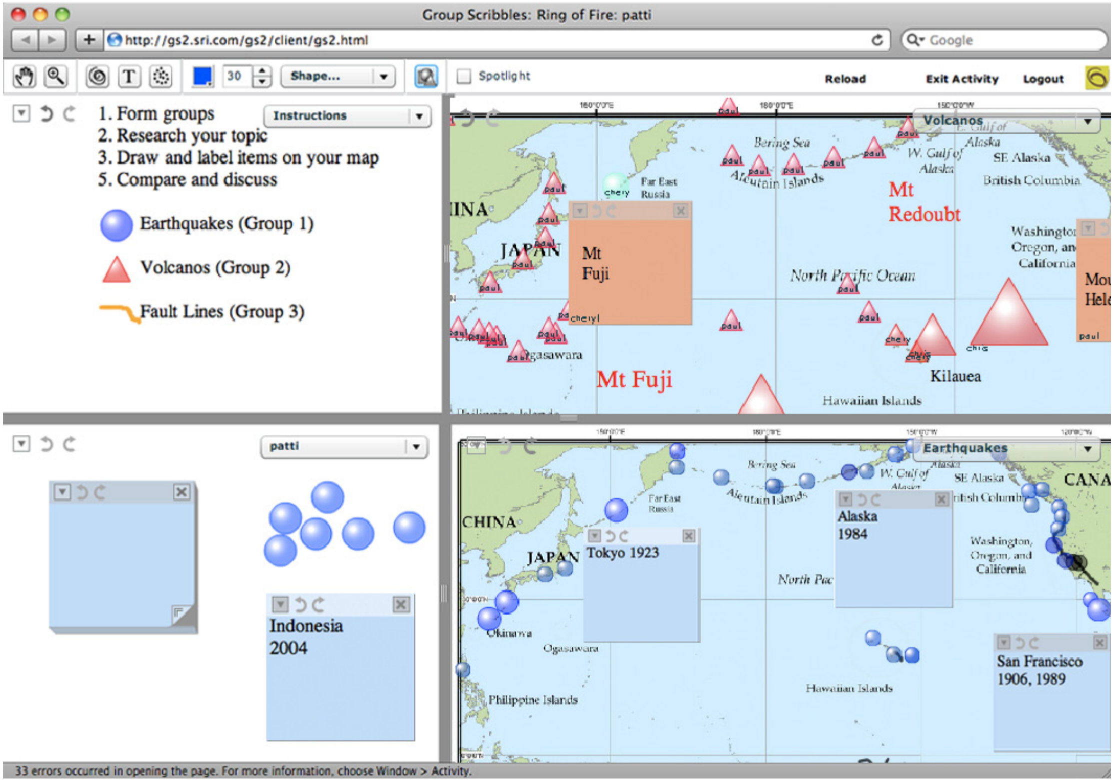


Figure 1. Screen Shot of Group Scribbles Boards.

A Tale of Two Projects

Group Scribbles is being used by the Grupo de Sistemas Inteligentes y Cooperativos (GSIC) research group at the University of Valladolid and in the Contingent Pedagogies project at SRI International. Since 2006, both groups have been working with Group Scribbles in several research efforts. Albeit independent, both groups were aware of each other's work as part of the tight research community around Group Scribbles. Initially, both institutions focused on the potential for Group Scribbles to support improvisation and social coordination in the classroom; subsequently, pedagogical patterns have taken on a more important role in the projects' work.

Group Scribbles is a participation-oriented network technology that supports collaborative activities using text, sketches, and images (Figure 1). The metaphor is based on common physical artifacts from the classroom: adhesive notes, whiteboards, pens and markers. Participants can scribble contributions on notes and post them anonymously in a shared public space that becomes the object of discussion. Teachers can quickly configure spaces for a short-term collaborative or group activity and, as the activity unfolds, alter the configuration on the fly and create new public boards to support multiple spaces for small groups to work.

GSIC Experiences with Group Scribbles

The GSIC research followed a bottom-up approach driven by a case study method (Stake, 2005). The research took place in five classrooms of a primary school (with 18–25 students each, ages 6–8) in Spain with eight teachers who had varying levels of teaching experience and ICT training. The researchers spent 2 years working closely with the teachers to help them orchestrate their activities using Group Scribbles and other ICT tools (e.g., digital whiteboards, tablet PCs) that were already in their classrooms. All the designed and enacted Group Scribbles activities were related to the teachers' usual curricula.

In the first stage of the research, Group Scribbles was introduced in a brief training session with teachers. Then, during activity design sessions, the researchers helped transform teachers' activity ideas into Group Scribbles activities. Thirty-one enactments of those activities were observed and analyzed. Additional information on the school and teacher context was gathered through three semi-structured interviews and a focus group with teachers. All sessions were audio-recorded, and observation notes were taken by at least two researchers for triangulation. See Villagrá-Sobrino & Prieto (2011) for more on the enactments analysis process.

The main result of this first stage of research was that the role of improvisation in teachers' practice was minimal, even when teachers used a tool like Group Scribbles that supports improvisation (Roschelle et al., 2007). They also acknowledged the importance of emergent occurrences in classroom enactment:

> T2: Maybe during the enactment of an activity many things can happen, which I have not prepared, but I think that it is important that teachers design their activities. [Teachers' focus group, 2009/03/17, translated from Spanish]

Further, teachers tended to design high-level tasks but their enactments had more small-scale, patterned kinds of improvisation. The activity patterns extracted from actual teaching were dubbed *routines* (both design routines and enactment routines) to differentiate them from researcher-specified *patterns*. Figure 2 (A) shows an activity design from a teacher's notebook. Figure 2 (B) shows the activity enactment analysis, reflecting its phases (distinct portions of activity enactment, often traceable to a design routine) and enactment routines (recurrent teacher moves present in the enactment) observed during each phase. The activity design (in bold) was completed by phases and routines that emerged during the enactment (e.g., R4a—"Disallow tool usage").

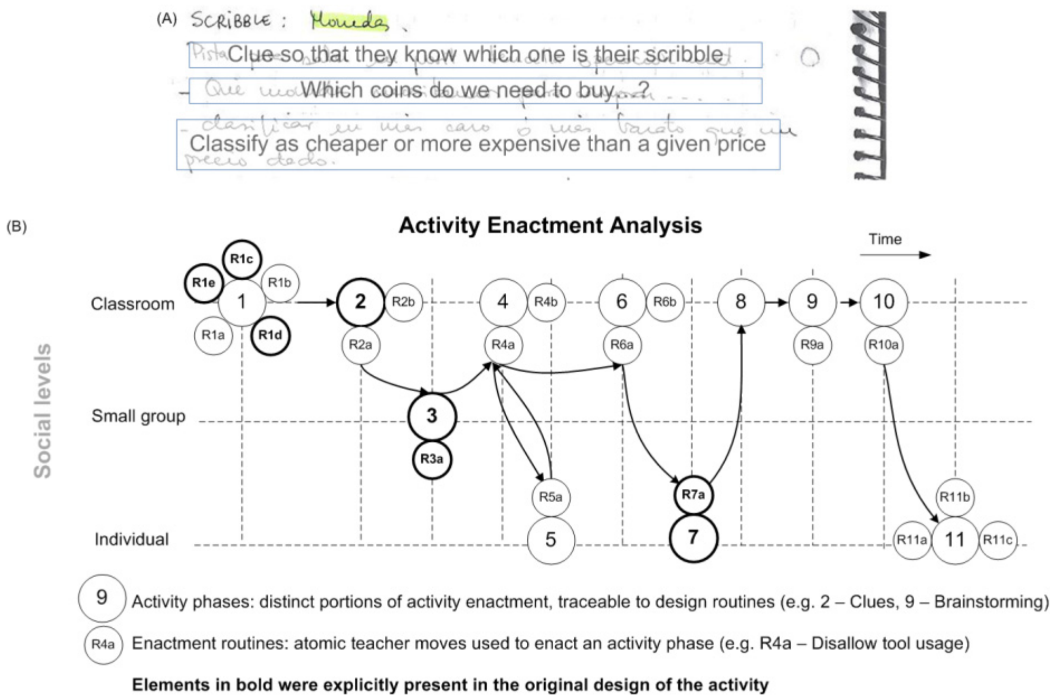


Figure 2. (A) Design, (B) Enactment Analysis of one Activity. Adapted from Prieto et al. (2010).

The goal of the second stage of this research was to foster collaborative learning practices among the teachers through different forms of professional development. A 2-hour training session on collaborative techniques centered on the use of collaborative learning flow patterns (CLFPs, see Hernández et al., 2010). As illustrated in the following excerpts from the session, the macro-level CLFPs presented were considered by teachers as too separated from their classroom contexts and practice, and thus not easily re-contextualizable:

> T3: It is very difficult to put into practice a role-play and a think-pair-share [names of CLFPs] with children in the first grade (6–7 years). [CLFPs training session, 2009/11/10, translated from Spanish]

T4: I would be hard-pressed to say anything I do that resembles these techniques [the CLFPs] [CLFPs training session, 2009/11/10, translated from Spanish]

The researchers speculated that routines elicited from teachers' real practice (such as the ones in Figure 2 above) might be a better entry point for teachers to design activities and to promote their reflection on the enactment (Prieto et al., 2010). A 2-hour professional development workshop was carried out with nine teachers from the same school (four teachers who were part of the previous research effort and five new to the research). After a brief presentation of the design and enactment routines from their practice, teachers were able to design and role-play an enactment of a complex Group Scribbles activity in 10 minutes. Moreover, a survey taken just after the workshop showed that the routines were familiar to them and that they appreciated their usefulness.

(To the question: Have you ever used these [design] routines? Which ones?)
T5: Yes, almost all: brainstorming, classification, ordering, where is on the image, Etc. [Survey after routines workshop, 2010/06/25, translated from Spanish]

(To the question: Did the [design] routines help you in enriching the design? Why?)
T6: Yes, because they bring new ideas about how to work on the same contents in different ways. [Survey after routines workshop, 2010/06/25, translated from Spanish]

The Contingent Pedagogies Project

The aim of the Contingent Pedagogies project is to improve student science learning by integrating assessment activities into a widely used Earth systems science curriculum, *Investigating Earth Systems* (IES), to create a comprehensive curricular activity system (Roschelle, Knudsen, & Hegedus, 2009). The Contingent Pedagogies project developed activities called interactive formative assessments (IFAs) for IES that use classroom network technology, clickers and Group Scribbles. The IFAs specify questions for teachers to pose; how network technology will be used to support collection, aggregation, and display of data; and how teachers can use assessment information to organize instruction. Pedagogical patterns successful in prior research for promoting individual and group learning served as templates for designing the IFAs (DeBarger et al., 2010).

In the first stage of research, five sixth-grade teachers who had experience implementing the IES curriculum were introduced to the network technology and pedagogical patterns in a professional development workshop. The teachers then worked together in small groups with an assessment researcher, curriculum developer, and subject matter expert to develop activities. After the workshop, the project team developed additional IFAs and worked with the five teachers to pilot-test the activities in their classrooms in Colorado. Each teacher taught at least three classroom sections in Earth science with 25–35 students each. The team engaged the teachers in monthly 1.5-hour teleconferences in which technology issues that teachers encountered were addressed, teachers reported on an activity that they implemented with students, and the team shared tips related to effective use of the technology. Technological support was also provided through web conferencing.

During the school year, Contingent Pedagogies researchers observed 12 classroom sections where teachers used Group Scribbles activities. Using semi-structured observation protocols, observers recorded the focal science topic and described teachers' and students' interactions. At the end of the class, observers recorded summaries of the class, including instructions teachers gave, teacher support for student engagement, variations in students' responses to the activity, questions posed and nature of the responses, breakdowns in the flow and management of activities, technology use, and communication of science content by the teacher. At the end of the year, an online survey was fielded to obtain a comprehensive view of all the activities and patterns the teachers attempted to implement and to identify activities and patterns that were more and less valuable or usable from the teacher's perspective. Detailed findings from this research appear in Penuel et al. (2010).

Both the survey and the observations indicated that the teachers could enact the Group Scribbles activities and patterns and that they did so many times during the 5 or so months when they were using the IES units. On average, each teacher used Group Scribbles activities six times in their classroom. Teachers reported that they had implemented most, but not all, of the activities developed by the team and that they had also created several of their own Group Scribbles activities based on the pedagogical patterns provided.

Results from the first stage of research suggested promising levels of adoption, but also highlighted challenges to implementation and the need for a broader set of tools to improve the quality of enactment of patterns in the classroom. The teachers felt that the patterns advanced the goals of enhancing communication, motivation, and feedback and that the IFAs helped students learn high-level skills. But they also experienced many tensions in classroom management, such as technical issues sidetracking lesson flow, figuring out the "right amount of time" to allow students to answer questions, and keeping students on task during group work.

T7: I want to have more training activities for my students to learn about it before they have to use it. I felt that was the hardest part this year; they were so enthralled to find out things that it was tough to manage. [Teacher teleconference, 2010/1/21]

> T8: The only issue for me is when they're doing the wrong things—writing on each other's boards, or writing a scribble that says "hi john". I take away their computer after a warning to stop doing that. [Teacher teleconference, 2010/3/25]

Further, the quality of the student participation and the contingent teaching observed were limited. Observers noted that teachers often asked students to explain their ideas, but teachers did most of the intellectual work of building on and connecting ideas and rarely engaged students in discussion of one another's ideas. Instead, teachers addressed issues by restating correct ideas or explaining why certain answers were incorrect. For example, after trying several times to explain a scientific idea related to divergent plate boundaries, one teacher asked students to "show by giving thumbs up, to the side, or down" the level of understanding they felt they had achieved. For those who displayed a thumb down (still confused) or to the side (not completely sure), the teacher told them that they could return to the classroom during lunch for further review.

Finally, the researchers found little evidence of teachers changing the direction of lessons to address clearly problematic ideas. Project staff had planned to introduce a set of contingent activities to teachers in a field trial the next year but concluded that providing contingent activities was insufficient; the teachers needed a broader suite of tools to improve the quality of enactment of patterns in classrooms. To support a more dialogic style in such enactment (O'Connor & Michaels, 2007), the project team developed a set of classroom norms for participation, discourse moves for discussion, and decision rules for contingent teaching, all of which are being implemented and studied with 15 teachers in the current (third year) of the project. Norms set expectations about how students will participate in discussion and establish a classroom community with a shared purpose of making sense of scientific ideas and practices (e.g., everyone will reason and respond; challenge ideas, not identities). Discourse moves (e.g., inviting students to build on a classmate's idea or summarize a key idea from a discussion) can help shift responsibility for thinking to students. Decision rules provide teachers guidance on how to proceed on the basis of assessment information. For example, if the class is divided between two alternative explanations, a teacher might break the class into two groups and ask students from each to pose questions to the other group about their explanations. Preliminary indications suggest promising uptake by teachers of these dialogic supports. Weekly logs from teachers show high levels of uptake of both the norms and discourse moves, and teachers report that they helped them advance their goals for instruction in nearly all instances when they employed them in conjunction with network technologies (Penuel & DeBarger, 2011). In focus group discussions, teachers offered accounts for why:

> T9: I really like the norms and I think they set a good standard in the classroom of what do to, and I see the kids using it in everything they're doing... The one I like the most is explaining, it's really making the kids support their reasoning. [Teacher teleconference, 2010/10/27]
>
> T10: I agree, it has really set a good tone for the kids... There are a lot of times when something is happening in class and I can refer back to those norms. Like the one norm about "it's okay to be wrong" based on your current understanding; that has come up on several occasions... We look at when you took that test that was what your understanding was... but now that you have this new understanding it's different. [Teacher teleconference, 2010/10/27]

Bridging the Gaps: A Challenge for Changing Everyday Practice

The findings and evolution of these two projects exemplify a common problem of CSCL research that tries to influence classroom practice in authentic settings: how to make the results of past research (often in the form of de-contextualized theories and principles or reified into new technological tools) available to practitioners in a way that they can appropriate them—and, moreover, how to do it for practitioners who are not especially gifted or motivated or are not experts (Dillenbourg, 2009). Studies into the sustainability and scalability of research-based interventions (Fishman et al., 2004; Penuel et al., 2007) point to the importance of such factors as the professional development approach, its coherence with current reform ideas in the schools, and the challenges that teachers face in their daily practice. Slavin & Lake (2008) also highlight the effectiveness of programs that address changes in teacher practice (e.g. by creating architectures for collaborative learning).

The projects depicted in this paper explored pedagogical patterns as a professional development approach to help address everyday teaching challenges. The pattern approach offers several advantages: it serves as a means of communication between researchers and practitioners (and also among practitioners), it offers practitioners a number of building blocks that can be creatively combined into new solutions, and finally, it is suitable for nonexperts because of its problem orientation. However, the evidence from the two projects described above shows that its application to everyday teaching practice is not without potential limitations. The findings of both research groups highlight two important tensions or gaps that often arise when researchers try to influence everyday practice in an authentic setting. These tensions are represented graphically in Figure 3.

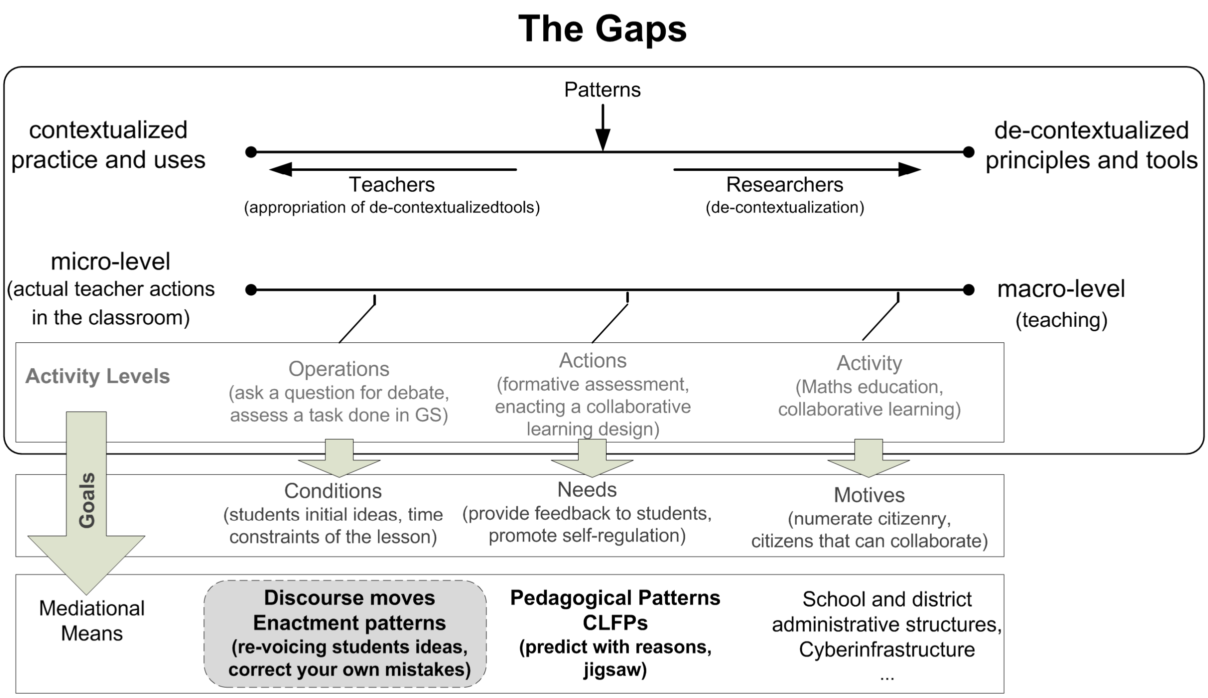


Figure 3. Gaps in the Contextualization and Mediational Support of Teaching. The focus of the research presented here is highlighted with a grey background (bottom-left).

The first of these tensions appears between researchers' efforts to de-contextualize empirical data to obtain widely applicable principles and teachers' application of those principles to concrete situations, which can be seen as an act of re-contextualization (Goodwin & Duranti, 1992). Teaching practice, and especially innovative teaching practice, can be seen as the appropriation of the de-contextualized tools such as curriculum materials, classroom management techniques, as well as resources provided by researchers (e.g. theories, patterns, or even technological tools such as Group Scribbles). In this sense, pedagogical patterns provide de-contextualized advice on how to attain certain pedagogical goals. However, as noted by Winters & Mor (2009), dealing with de-contextualized tools can be difficult for teachers, even if they contain more elements of context than an abstract theory (e.g. they assume a certain kind of classroom, or they are provided along with a short narrative example of their application to other contexts). Instructional moves and design and enactment routines elicited from actual teaching practice are also examples of this de-contextualization effort, but they originate from a different source than theory. Having more elements of a familiar context present in these patterns (e.g. assuming usage of the Group Scribbles tool and a whiteboard, or assuming a specific outcome of a previous task) enhances their mirroring properties and makes them more actionable (i.e. teachers recognize them as actions that they normally take in the classroom or may take in an easily recognizable situation).

There is a second tension or gap between the macro-level designs and plans for instruction (provided by researchers, developed by teachers, or co-designed) and the emergent micro-level enactment of those plans by a specific teacher in a classroom. Such plans are incomplete by their very nature, since any representation of a practice is a simplification. Even if plans are designed by teachers thinking about their specific classroom context, plans cannot take into account all emergent occurrences or accurately predict students' notions and their evolution. If we look at this gap from the point of view of socio-cultural activity theory (Engeström, 1987), lesson plans and pedagogical patterns at the macro-level (e.g. the CLFPs mentioned above) provide mediational tools for teachers at the action level that respond to needs such as providing feedback to students or promoting self-regulation. But even with that scaffolding, teachers must still make decisions on how to enact the plans using specific instructional moves in their classroom context. These instructional moves (which correspond to operations in activity theory terminology) can be highly routinized and often vary based on teacher style. By also providing scaffolding at the operation level (e.g. Contingent Pedagogies' discourse moves, or GSIC's enactment patterns), more coherent pedagogical strategies can be enacted by teachers. Having a set of atomic, actionable patterns that are easy to call forth, tweak and recombine can empower teachers to creatively design and enact activities according to the theories and design principles of CSCL research (Hernández et al., 2010) and dialogic teaching research (Wells & Mejia-Arauz, 2006; O'Connor & Michaels, 2007).

Moreover, this combination of patterns of different granularities is supported by Alexander's concept of a pattern language (Alexander et al., 1977), that is, a set of related patterns that provide increasing detail on how to implement the higher-granularity patterns. In our case, norms, rules, moves, and routines can be seen as tools for goal-directed action that ideally become operationalized in ways that support teachers' enactment of collaborative, dialogic activities. They also help teachers in specifying further innovations using technology in ways that can enhance their implementation (Cohen & Ball, 1999).

Conclusions and Future Work

The pedagogical patterns approach has been motivated by the need to exchange knowledge and good practices between research literature and the real world, as a way to support practitioners and as a means of communication among various stakeholders (e.g., teachers and education or technology researchers).

However, researchers who use patterns in CSCL still encounter challenges in changing everyday classroom practice. This paper has presented two CSCL projects that tried to take contingent, adaptive teaching with flexible network technology (Group Scribbles) to the real world through the use of patterns. Despite the differences in school context or even the overall research approach, researchers on both projects independently identified and analyzed several common issues that must be addressed. One is the gap between the de-contextualized theories and tools that researchers often produce and teachers' need to provide ad-hoc practice in their classroom situations (which can be seen as an act of re-contextualization). Another is the gap between the macro-level advice (e.g. in the form of pedagogical patterns or lesson plans) that is often given to teachers and the micro-level decisions and actions that teachers must take in their particular contexts. In both cases, the use of more atomic, actionable teacher moves (coming from real practice and derived from literature) has showed promising results. The use of practice-derived enactment patterns, classroom norms, and decision rules seems to complement the advantages of macro-level pedagogical patterns and to enable a wider adoption and change of daily practice. Thus, we posit the combination of both kinds of patterns as a coherent mediational strategy for teachers to produce contextual, pedagogically-sound uses of technology, making the most of its affordances for enhancing teaching and learning.

Further research is needed to accumulate more evidence in favor of or against this proposal for an effective use of pedagogical patterns in CSCL. Our research teams in the United States and Spain intend to build on the findings of both projects and thus tackle a common problem in CSCL—the low reusability of knowledge. For example, we plan to explore the application of patterns (including routines, moves, rules, and norms) across both projects and analyze teacher practice in the U.S. context to identify additional routines that can be used in professional development workshops in conjunction with research-driven patterns. By addressing these issues in two contexts, we hope to generate insights into how to prepare teachers to become more proficient in orchestrating collaborative discussions and in enacting contingent teaching with ICT in real-world classrooms.

References

Alexander, R. (2008) *Towards dialogic teaching: Rethinking classroom talk*, 4th edition. York, UK: Dialogos.

Alexander, C., Ishikawa, S., & Silverstein, M. (1977). *A pattern language: Towns, buildings, construction*. New York, NY: Oxford University Press.

Anderson, R., Anderson, R., Davis, P., Linnell, N., Prince, P., Razmov, V., & Videon, F. (2007). Classroom Presenter: Enhancing interactive education with digital ink. *IEEE Computer*, *9*, 56-61.

Beatty, I., Gerace, W., Leonard, W., & Dufresne, R. (2006). Designing effective questions for classroom response system teaching. *American Journal of Physics, 74*(1), 31-39.

Cohen, D. K., & Ball, D. L. (1999). *Instruction, capacity, and improvement*. Philadelphia: Consortium for Policy Research in Education.

Conole, G., McAndrew, P., & Dimitriadis, Y. (in press). The role of CSCL pedagogical patterns as mediating artifacts for repurposing Open Educational Resources. In F. Pozzi & D. Persico (Eds.), *Techniques for fostering collaboration in online learning communities: Theoretical and practical perspectives*. Hershey, PA: IGI Global.

DeBarger, A., Penuel, W., Harris, C., & Schank, P. (2010). Teaching routines to enhance collaboration using classroom network technology. In F. Pozzi & D. Persico (Eds.), *Techniques for fostering collaboration in online learning communities* (pp. 222-244). Hershey, PA: IGI Global.

DiGiano, C., Yarnall, L., Patton, C., Roschelle, J., Tatar, D., & Manley, M. (2003). Collaboration design patterns: Conceptual tools for planning for the wireless classroom. *Journal of Computer-Assisted Learning, 19*(3), 284-297.

Dillenbourg, P. (2009). Exploring neglected planes: social signals and class orchestration. *Proceedings of the 9th International Conference on Computer Supported Collaborative Learning - Volume 2* (pp. 6-7). New Brunswick, NJ: International Society of the Learning Sciences.

Dimitriadis, Y., Asensio, J., Hernandez, D., Roschelle, J., Brecht, J., Tatar, D., Chaudhury, S., DiGiano, C., & Patton, C. (2007). From socially-mediated to technology-mediated coordination: A study of design tensions using Group Scribbles. *Proceedings of the 8th International Conference on Computer Supported Collaborative Learning* (pp. 184-186). New Brunswick, NJ: International Society of the Learning Sciences.

Engeström, Y. (1987). *Learning by expanding: an activity-theoretical approach to developmental research.* Helsinki, Finland: Orienta-Konsultit.
Fishman, B., Marx, R., Blumenfeld, P., Krajcik, J., & Soloway, E. (2004). Creating a framework for research on systemic technology innovations. *The Journal of the Learning Sciences, 13(1)*, 43-76.
Goodwin, C., & Duranti, A. (1992). Rethinking context: An introduction. In A. Duranti & C. Goodwin (Eds.), *Rethinking context: Language as an interactive phenomenon* (pp. 1-42). New York, NY: Cambridge University Press.
Hernández, D., Asensio, J. I., Dimitriadis, Y., & Villasclaras, E. D. (2010). Pattern languages for generating CSCL scripts: From a conceptual model to the design of a real situation. In P. Goodyear & S. Retalis (Eds.), *E-learning, design patterns and pattern languages* (pp. 46-64). Rotterdam, The Netherlands: Sense Publishers.
Kennewell, S., Tanner, H., Jones, S., & Beauchamp, G. (2008). Analysing the use of interactive technology to implement interactive teaching. *Journal of Computer Assisted Learning, 24* (1), 61-73.
Looi, C. K., Chen, W. & Ng, F. (2010). Collaborative activities enabled by Group Scribbles (GS): An exploratory study of learning effectiveness. *Computers & Education, 54* (1), 14-26.
Nystrand, M., & Gamoran, A. (1991). Instructional discourse, student engagement, and literature achievement. *Research in the Teaching of English, 25*(3) 261-290.
O'Connor, C., & Michaels, S. (2007). When is dialogue "dialogic"? *Human Development, 50*, 275-285.
Penuel, W. R., & DeBarger, A. H. (2011, April). *Supporting teacher learning to improve classroom assessment in science.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.
Penuel, W., Roschelle, J., & Abrahamson, A. (2005). Research on classroom networks for whole-class activities. *Proceedings of the IEEE International Workshop on Wireless and Mobile Technologies in Education* (pp. 222-229). Los Alamitos, CA: IEEE.
Penuel, W., Fishman, B., Yamaguchi, R., & Gallagher, L. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal, 44(4)*, 921-958.
Penuel, W. R., Beatty, I. D., Remold, J., Harris, C. J., Bienkowski, M., & DeBarger, A. H. (2010). *Pedagogical patterns to support interactive formative assessment with classroom response systems.* Menlo Park, CA: SRI International.
Prieto, L. P., Villagrá-Sobrino, S., Dimitriadis, Y., Jorrín-Abellán, I., Martínez-Monés, A., Anguita-Martínez, R. (2010). Recurrent routines in the classroom madness: pushing patterns past the design phase. *Proceedings of the 7th International Conference on Networked Learning* (pp. 499-507). Lancaster, UK: Lancaster University.
Roschelle, J., Knudsen, J., & Hegedus, S. (2009). From new technological infrastructures to curricular activity systems: Advanced designs for teaching and learning. In Jacobson, M. J., & Reimann, P. (Eds.). Designs for Learning Environments of the Future: International Perspectives from the Learning Sciences. New York: Springer.
Roschelle, J., Tatar, D., Chaudbury, S., Dimitriadis, Y., Patton, C., & DiGiano, C. (2007). Ink, improvisation, and interactive engagement: learning with tablets. *IEEE Computer, 40(9)*, 42-48.
Slavin, R.E. & Lake, C. (2008). Effective programs in elementary mathematics: A best-evidence synthesis. *Review of Educational Research, 78* (3), 427-515.
Stake, R. (2005). Qualitative case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (pp. 443-466). Thousand Oaks, CA: Sage.
Villagrá-Sobrino, S., & Prieto, L. P. (2011). *The making of a routine catalogue*. Valladolid, Spain: University of Valladolid. Available online at http://www.gsic.uva.es/~lprisan/Villagra2011_RoutineAnalysis.pdf
Wells, G., & Mejia-Arauz, R. (2006). Dialogue in the classroom. *Journal of the Learning Sciences, 15(3)*, 379-428.
Winters, N., & Mor, Y. (2009). Dealing with abstraction: Case study generalisation as a method for eliciting design patterns. *Computers in Human Behavior, 25(5)*, 1079-1088.

Acknowledgments
This material is based on work supported by the National Science Foundation under grant numbers 0713711 and 0822314, the Spanish Ministry of Education and Science project TIN2008-03023/TSI, and Autonomous Government of Castilla and Leon (Spain) projects VA107A08 and VA106A08. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the funding agencies. The authors thank the teachers and students of the CEIP Ana de Austria school and Denver Public Schools for their participation, support, and significant contributions to this research.

Learning Design as a Medium for Scaffolding Teacher Learning and Collaboration

Nancy Law, University of Hong Kong, Pokfulam Road, Hong Kong, nlaw@hku.hk
Diana Laurillard, London Knowledge Lab, Institute of Education, 23-29 Emerald St, London WC1N 3QS, d.laurillard@ioe.ac.uk
Yeung Lee, University of Hong Kong, Pokfulam Road, Hong Kong, yeunglee@hkucc.hku.hk

Abstract: Teachers implementing new curricula targeting the development of 21st century skills face the challenge of learning how to design learning experiences that align with the intended pedagogical approach. Curriculum reform initiatives expect that development will take place at least in part through teacher collaboration and sharing of curriculum resources through digital repositories. However, curriculum artifacts have not so far proved to be an effective medium for teacher collaboration. This paper argues that the teaching profession suffers from not having a shared professional language to describe and communicate their design ideas, and demonstrates that the provision of a design language and format that highlights the multilevel, multi-faceted aspects of a design helps to scaffold teacher discourse to deeper pedagogical discussions. On the basis of the findings from a cross-cultural collaboration, we set out the implications for computer-supported collaborative learning for teachers, to enable more effective teacher learning and collaboration.

Introduction

There have been increasing pressures from businesses and governments to move education from a focus on competence in specific knowledge and skills to developing students' information literacy, problem solving, collaboration and communication skills—abilities that are generally referred to as 21st century skills (Partnership for 21st Century Skills, 2003.; UNESCO, 2008). The demand is for education to produce graduates who are capable of knowledge creation and innovation, which requires changes in the goals, curricula and processes of schooling (e.g. Kozma, 2008; UNESCO, 2008; CERI, 2001). In response to such pressures, education policy documents in many countries show a strong orientation towards more student centered modes of learning and pedagogies that encourage students' active engagement in collaborative inquiry and the use of ICT to support collaboration and inquiry (Pelgrum & Law, 2003; Plomp, Anderson, Law & Quale, 2009). These changes can only be realized through deep changes in teachers' knowledge, beliefs and practice. It is thus not surprising that many major curriculum reform initiatives are accompanied by significant investment to provide support for teacher learning and curriculum development, as well as teacher collaboration and sharing of curriculum resources through digital repositories.

In Hong Kong (and possibly in many other countries), curriculum and professional development projects are expected to deliver, as an important outcome, curriculum artifacts generated by teachers and/or researchers that can be used by other teachers in implementing new curriculum and pedagogy. Hence curriculum resources act as a kind of defacto medium for teacher collaboration and professional learning. On the other hand, it has been observed that the adoption of open education resources is much lower than the amount of materials that have been shared (1). So we have a somewhat contradictory observation that teachers complain about lack of suitable resources for implementing new goals and approaches in the school curriculum, while curriculum resources specifically designed for teachers, including those by teachers, are not being adopted. Furthermore, even in cases where the curriculum artifacts are adopted by another teacher, it is often the case that the use is different from how it was initially intended or designed for in the first instance (Laurillard and McAndrew 2003). Thus, curriculum artifacts have not proved to be an effective medium for teacher sharing, nor have they been an effective medium for communicating pedagogy.

The biggest challenge to teachers in implementing new curricula targeting the development of 21st century skills is in learning how to design learning experiences that align with the intended pedagogical approach. There are three different aspects of curriculum design on which teachers typically collaborate: theories about learning and pedagogy, teaching and learning activities, and resources for teaching and learning. While these types of teacher interactions are useful and important, they are not sufficient for negotiating the full complexity of the learning design process. To ensure that learners are provided with learning opportunities to achieve the targeted outcomes, teachers have to engage in design work at several different but interconnected levels: the curriculum unit (or module), the sessions (or lessons) comprising the unit, teaching and learning activities before, during and after a session, and the resources and tools to be used within a session. At each of the design levels, the teacher has to give due consideration to the learning goals (content, skills and/or attitude), pedagogical approach to be adopted, the nature of the student engagement needed, what student-generated content needs to be specified, the rubric or criteria for assessment, and the kind of feedback that will be provided to the learner. In addition, the teacher needs to ensure that the different levels and facets of the design are properly aligned to provide a coherent set of appropriate learning experiences (Biggs 2003).

In this paper, we put forward the view that the teaching profession suffers from not having a shared professional language to adequately describe and communicate these design issues and decisions, and demonstrate that the provision of a design language and format that highlights the multilevel, multi-faceted aspects of a design helps to scaffold teacher discourse to deeper pedagogical discussions. We will also discuss, on the basis of the findings, implications for the design of technology support for learning design that will be able to support more effective teacher learning and collaboration.

Context of the Study

The present study is conducted in the context of a university-school partnership project, titled Learning 2.0 (2), which aims to design a curriculum and assessment platform to be used to support the teaching of Liberal Studies, a new, compulsory subject in the Hong Kong New Senior Secondary School Curriculum launched in September 2009 (CDC & HKEAA, 2007). This subject was introduced by the Education Bureau to address the issue of an over-emphasis on rote-learning. The subject focuses on fostering students' lifelong learning and inquiry skills. The curriculum specifies a number of key concepts such as social mobility, quality of life and globalization in the humanities, science and technology areas for students to develop a reasonable understanding through exploring issues emerging from contemporary and current themes and problems. There is no a set textbook or content specification, but the curriculum specifies *issue-enquiry* as the pedagogical approach of choice that teachers should adopt in teaching this subject. Furthermore, teachers are expected to play the role of curriculum and assessment designers to develop modules to implement this issue-enquiry approach for the achievement of the curriculum goals for this subject.

The *Liberal Studies Curriculum and Assessment Guide* (CDC & HKEAA, 2007, to be referred to as the C&A Guide) recommended some design principles on implementing the curriculum, with implications for how issue-enquiry modules should be designed. First of all, these modules should provide contexts for students to learn to integrate, apply, consolidate and broaden their foundational knowledge through engaging in in-depth inquiry and reflection on contemporary issues connected with the selected contexts. Students should be exposed to perspectives and concepts essential to the understanding of issues of human concern, and develop their independent and lifelong learning skills, values and attitudes, which could be transferred and applied to the understanding of new issues and contexts. According to the C&A Guide, each module should be organized around a central concept within one of three areas of study specified in the curriculum (e.g. Society and Culture). Each module should identify questions for inquiry related to key themes (e.g. quality of life, rule of law and socio-political participation) relevant to students' lives, and embody perennial issues (i.e. issues involving values that are important to students and society and underpinned much of the debates and conflicts surrounding the problem context).

The Learning 2.0 project is funded by the Hong Kong Quality Education Fund. An important component of the project is to develop a Moodle-based (3) online learning and assessment platform (iLAP) (4) that teachers can use to create online course-rooms for their students as an integral part of the teaching and learning activities in the Liberal Studies subject. Participating project schools identify key teachers to contribute to the design and development of iLAP, as well as the design, implementation and evaluation of the inquiry enquiry modules on this platform. Teachers involved in this project hold weekly meetings to discuss and share module ideas and designs (all meetings are audio-recorded, and are used for the analysis reported in this paper). They are also able to visit each other's actual course websites (referred to as course-rooms in the project) on the iLAP platform. The iLAP platform enables the teachers in this project to collaborate and share the full richness of their teaching designs in terms of activities and resources used as well as review the student-generated work in these course-rooms.

Nature of Teacher Discourse on Designs Presented in the iLAP Course-rooms

Teachers in the network hold weekly project meetings and one of the regular items on the agenda is to share design ideas and implementation experiences on the curriculum modules they have set up on iLAP. The teachers are generally very interested in sharing their work and discussing ways of improvement. On the other hand, the discussions were at a level that made it difficult to move the designs forward at a deeper pedagogical level. To give some concrete illustration of the focus of the discussions, the following is a list of the issues/suggestions raised in one of the teacher meetings on two of the modules designed and implemented by teachers in two different project schools in June and July of 2010 (these two modules will also be the focus of our further exploration on ways to enhance teacher sharing and collaboration later in this paper):

1. Recommend the use of the forum function in "group discussion" activities, as this will allow the discussion to be commented on and assessed by peers and by the teacher using the built-in assessment function.
2. Suggest getting students to work on the Wiki for consolidating what they have learnt through the module. The availability of the history function in Wiki will allow teachers to track students' contributions.

3. Suggest the use of Mahara on the iLAP platform for students to build a module folder
4. It may be better to provide every student with his or her own individual Wiki page.
5. The self-reflection tasks may be better done on paper.
6. Which will be better for student learning: online forum v.s. a face-to-face debate?
7. One teacher shared a generic rubric on critical thinking and suggested providing such for students' reference as an implementation of the "assessment for learning" idea.
8. Suggest getting students to answer questions from the C&A Guide as a after-lesson consolidation
9. Students in some schools may not have the motivation to participate in collaborating on the wiki or in self-reflection activity
10. The allocation of two weeks for the module limits the extent to which relevant iLAP tools/activities can be integrated into the module

It can be seen from the above list that the teachers were very engaged in sharing and discussing pedagogical design during these meetings, particularly in relation to the effectiveness of using different technological tools for specific learning activities (points 1 to 6 of the above list). Points 7 and 8 were related to assessment and learning activity design to enhance students' learning that could be applicable for the subject in general. Point 9 was about the match between the motivation (and ability) of students and a specific type of online activity. All these discussions were valuable, but were pitched at a very general level. Only point 10 was about the specific feature of the particular module under discussion. Moreover, none of these discussions were specific to designing issue-enquiry pedagogy as specified by the C&A Guide.

Comparing Learning Designs as Communicated through Artifacts in an Online Course-room

The project research team was somewhat disappointed about the lack of more holistic discussions about module level pedagogical design for issue-enquiry. In addition, the foci of the teachers' discussions for both of these modules were very similar despite important pedagogical differences between them as perceived by the research team. Further explorations reveal that the activity structures of these two modules were very similar as can be seen from Table 1, which presents a summary of the information available in the two respective course-rooms. Both identify a conceptual focus and generic inquiry skills as the learning goals and a current affair issue as the context for students' exploration. Both started with using media materials (videos from TV broadcasts and newspaper clips) to introduce the issue context and to stimulate student discussion. This was followed by group discussions guided by a worksheet for a focal student activity in the module—a role-play in Module A and a debate in Module B. Students were provided with a structured wiki to continue their group preparation at home. Finally, after the focal activity, students were asked to complete a reflection activity. The total class contact time was the same—five 35-minute periods. Module A has a double period in session 4. Sessions 2 and 3 in module B were double periods. Upon detailed inspection, there are differences at the level of specific activities such as the nature of the reflection task. It is apparent that these are the critical differences at the specific activity level that constituted each teacher's focus of attention on these designs.

The Structure of a Pedagogical Pattern

It can be argued that what is being presented in the online course-rooms are not adequate representations of teachers' curricular designs. The challenge to us is—what would be an adequate representation? We need a common format that is capable of revealing the similarities and critical differences in teachers' pedagogic designs.

A research project in the UK is developing a prototype for a 'learning design support environment for lecturers' (LDSE) (5) with the explicit intention of having an impact on teachers' practice in designing technology enhanced learning (TEL) by giving them the means to represent and share their design ideas. . In this study, the two projects are collaborating to test the extent to which the prototype meets the requirements of teachers who need help with making good use of technology, and whether it succeeds in promoting deeper pedagogical discussions and more collaborations in pedagogical design.

The requirements for a design pattern of any kind are that it must have a context, problem to be solved, and a way of solving the problem (Goodyear 2005). For a pattern that captures a good pedagogic design, the context is formal education, the problem is the learning outcome to be achieved, and the solution is the sequence of teaching-learning activities that have been found to succeed in that. Representing pedagogy succinctly and intelligibly is not easy. The current formats for representing learning design cover several parameters relevant to the teaching context, such as rationale, roles, group size, etc., but they disguise the complexity of the pedagogy by placing it in a single category such as 'sequence of activities' (Laurillard and Ljubojevic, in press). Alternatively, the narrative account of the learning design (6) is detailed, but difficult to generalize and customize to one's own context. The pedagogical pattern format that can scaffold productive pedagogical discussions, professional collaboration and development has to unpack the sequence of teaching and learning activities to expose the crucial features that makes a pedagogy work.

Table 1: A comparison of two issue-enquiry modules as communicated through the iLAP online course room.

	Module A	Module B
Module title	Fine-tuning of medium of instruction (MOI) policy in Hong Kong	Understanding post-'80s youth through the high-speed rail conflict
Learning goals	Understand the concept "quality of life" and learn generic skills of inquiry	Understand the concept "political participation" and learn generic skills of inquiry
Module context	The change in language policy by the Education Bureau in Hong Kong	Post-80s youth staging high profile demonstrations against the construction of high-speed rail through a remote village
Module level resources	Teacher generated handout containing definitions of key concepts, and list of key policy changes related to context	A group wiki on concepts for students to write down and share the most appropriate definitions of key concepts
Session 1	• Video 1—an interview (for teacher presentation) • Group wiki/worksheet 1 on video 1 (for use in groups discussion) • Video 2—an interview (for teacher presentation) • Student worksheet 2 on video 2(for individual work in class)	• Pre-session preparation—online discussion forum for students to discuss issues related to the context and to summarizes key views found in discussion in a group wiki • Handout 1 containing newspaper clips on context—teacher presentation • Supervised group discussion to complete group wiki/worksheet • Video 1—a televised public debate—teacher presentation • Supervised group discussion to complete group wiki/worksheet • Post-session review—students read Handout 2 with more newspaper clips and complete online quiz
Session 2	• Video 3—a TV feature on the policy change (for teacher presentation) • Student worksheet 3 on video 3 (for individual work in class) • Political cartoon (for teacher presentation) • Group wiki/worksheet 4 (for use in groups discussion)	• Pre-session preparation—students read Handout 3 (newspaper article by one post-80s activist) • Supervised group discussion to prepare for role play using group wiki/worksheet
Session 3	• Group wiki/worksheet 5 (used in group discussions to prepare for role play) • Four sets of newspaper clips – one for each stakeholder in the role play • Self-evaluation rubric for students to assess their own preparation	• Debate • Group concept mapping as a part of debriefing consolidation
Session 4	• Role-play • Evaluation rubric for assessing role-play • Student to submit reflection essay (online)	

Our current template for a 'pedagogical pattern' consists in the following categories of description:

- Title – to capture the point of interest;
- Learning outcome – to help the teacher decide if it is likely to fit their need;
- Summary – similar to a brief abstract;
- Sequence of teaching and learning activities (TLA) with the time needed for each – time on task is a critical aspect of pedagogic value;
- The category of TLA each one falls into – to assist analysis and comparison;
- A description of each TLA that reflects the core focus and critical aspect in implementation
- The assessment used to evaluate learner attainment
- Tools and resources needed – to promote the sharing of these across patterns, and to enhance them by embedding them in a pedagogical pattern
- The designer's reflections – to promote reflective use and revision of the pattern

A pedagogical pattern is developed initially from a specific instance, as that is where the effort of design is to be found, where a teacher is intent upon achieving the best outcome for the learners they know, in the topic they know. Once the teaching idea is articulated in a formal pattern description it is often easier for the teacher to see what is missing or what could be improved, so this can be a useful exercise in its own right. An example is given in Figure 1, using a pattern derived originally from the iCOPER project (7). This was developed for an initial teacher-training context, represented by the content details defined at the top: *classroom teaching, photographing key situations*, and *children's engagement*. Each of the content items is colored differently to highlight specific aspects important to the particular instance of the design. For example the pattern in Figure 1 can be replaced with new content items, e.g. *dental surgery, video recording*, and *patient care*, to adapt the pattern to a quite different professional development field (8). The essence of the pedagogy is captured in the unchanging text, making it possible for good teaching ideas to be shared more easily across discipline boundaries. The colored text in the activity sequence in Figure 1 indicates how the specific curriculum aspects of the design are taken care of in the TLA sequence. The task for our collaborative project is to test whether this approach works for teachers in schools, and whether they can use it to enhance the process of exchanging ideas on pedagogic design.

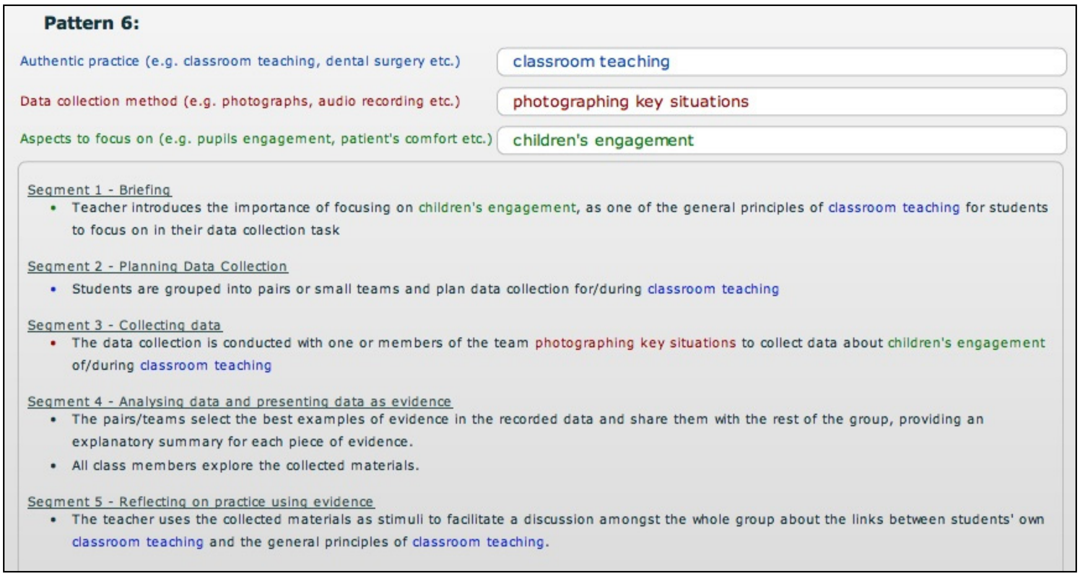


Figure 1. A 'Pedagogical Pattern' Instantiated with the User's Content for an Initial Teacher-training Course.

Comparing Learning Designs Represented on a Design Template

A preliminary pedagogical pattern was constructed to capture the essence of the features and pedagogical decisions in the design of issue-enquiry modules for the present study. We identified five critical elements in the pedagogical pattern—conceptual learning outcomes targeted (e.g. quality of life, political participation), generic skills and attitude outcomes targeted (e.g. able to identify key points of debate, be open and listen to different opinions), the issue context for the module (e.g. the Fine-tuning of the MOI policy in Hong Kong, construction of the high-speed rail through a remote village), teacher-provided content (e.g. news media, assessment rubrics), and student-generated content (e.g. wiki, online forum). Figure 2 presents a design template we developed for use in this study to represent the designs for Modules A and B described above.

The pedagogical patterns as constructed using this template highlighted important differences between the two modules. For Module A, there was no activity that explicitly addresses the conceptual learning outcomes; generic skills was addressed through the assessment rubrics, focusing on students' presentation and communication skills); and most of the activities were centered around the specific context—the impact of the policy change on different stakeholders. For Module B, we see a gradually changing focus of the activities from the specific context (construction of high-speed rail) to identification of the focus of the debate to linking the debate to some key concepts related to political participation. There is a strong focus on developing students' conceptual understanding through teacher-led discussions as well as the tasks designed for group discussion and unsupervised individual student work.

Effect of Learning Design Representation on Teachers' Professional Discourse

In September 2010, six new schools joined the project as a second phase to scale up the innovation. Modules A and B were presented to all of the participating teachers, most of whom are new to the project. In addition to the

Descriptors for use at module level:
- Module content goals: specifying the **concepts**
- Module generic skills & attitude goals: **skills, attitude**
- Issues forming the context of the module: **context**
- General references/resources at module or course levels: **resources**
- Students' individual or group work, supervised and unsupervised: **student-generated content**
- Module length: N sessions, each T-min long
- Venue requirements:
- Teaching support

Descriptors for use at session level:
- Specific learning goals:
  - Content goals
  - Skills goals
  - Attitude goals
- Issues to be explored
- Stages of inquiry
  1. Defining problem of inquiry
  2. Accessing and interpreting and information
  3. Analysis and organization of information
  4. Collaborative co-construction of knowledge
- Pedagogical approach
  - Teacher-directed
  - Student-centered
  - Student-directed
- Teacher activities
  - Teacher presentation—mainly on content
  - Teacher-led discussion
  - Teacher guidance—on procedures and administration
  - Teacher facilitation—on process and requirements
- Student activities
  - Group work – supervised
  - Group work – unsupervised
  - Peer review – supervised
  - Peer review – unsupervised
  - Well-structured group work, e.g. role play
  - Individual supervised work
  - Individual unsupervised work
- Resources
  - Teacher generated content—context-related
  - Teacher generated content—conceptually-focused
  - Teacher generated content—assessment-related, e.g. rubric
  - Student generated content—closed-ended
  - Student generated content—open-ended
- Venue requirement
- Teaching support requirement

Figure 2. A Design Template Used for Generating the Pedagogical Patterns for Modules A and B.

iLAP course room, the teachers were presented with the pedagogical patterns for the two modules and asked to provide comments. From the pedagogical patterns, the teachers very quickly identified the difference in the nature of the content focus for the two modules. A very lively discussion ensued that covered issues not observed in previous discussions. The key issues raised by the teachers include:

1. Module A could be improved through more appropriate consolidation activities such as asking students to write a short essay that explicitly reflect on the link between the MOI policy with the key concepts related to quality of life that the module wants to target.
2. Students [in module A] should be guided to demonstrate their understanding of the targeted concepts through their discussions on the contextual issues.
3. Module B is able to achieve a better alignment with the key conceptual outcomes targeted as students were guided to identify the conceptual issues underpinning the conflicts rather than simply identifying the view points on the contextual issues [as in Module A].
4. Module B has a context/scenario-based design in that students are introduced to the context and issues before the content. This design is better as students will first develop a stronger empathy with the contextual issues, making the students feel that the conceptual issues are relevant to their everyday life.
5. Opposing viewpoint to the above was also expressed—Module A adopts a deductive approach of introducing the concepts before the context, which could be clearer and more explicit for students to follow, and is hence preferable to the inductive approach of first introducing the context.
6. Some teachers think that both inductive and deductive approaches could work, but it is important for the teacher to be fully aware and be explicit about the approach that is being adopted.
7. The current issues selected as the context of enquiry should not be the focus of enquiry in a Liberal Studies module—it only serves the purpose of stimulating explorations of the targeted conceptual outcomes.
8. Current issues easily become outdated and lose their relevance. There should be continuity from one module to the next—we should find ways to check and ensure that students are able to apply the concepts learnt in one module to the analysis of issues in another, different context. This is the core design idea of the Liberal Studies curriculum.
9. There is no need to address every relevant concept, skill and attitude in one lesson or in a single module. Sometimes it may be better if the teachers just focus on one focal learning outcome in one lesson.
10. Teachers should pay close attention to the questions in the C&A Guide in designing learning tasks such as student discussions to ensure a better alignment with the targeted learning goals.

There are two prominent features in this discourse. First of all, there is a strong focus on the alignment between the targeted learning outcomes and the design at individual TLA and module levels. Secondly, the discourse encompasses design considerations at activity, module and curriculum levels. Points 1 to 3 focus on how the alignment could be improved through changes in the selection or implementation of individual learning activities. Points 4 to 6 concern the sequencing of activities and the implicit pedagogical model underpinning them, which is a module level design issue. Point 7 addresses a core design concern in issue-enquiry pedagogy—the role of context in a module. Point 8 goes beyond a single module to discuss the connection between modules within the context of the whole curriculum. Point 9 is a design concept that could be applied to the whole curriculum and point 10 refers explicitly to how the C&A Guide could be used in the process of module design. This discourse is clearly much richer in terms of pedagogical design considerations compared to the discussions on the same modules when the pedagogical patterns were not available. This is an indication that an appropriately structured learning design representation can potentially make a difference to teachers' learning and collaboration discourse in pedagogical design.

We were also interested in teachers' views of the usefulness of the pedagogical pattern template in supporting teacher sharing and collaboration. Interestingly, the response was less than positive. The general feedback is that the template is not user-friendly, and the many colors (actually five was used, as presented in Figure 2) in the text were somewhat overwhelming. The teachers were particularly concerned and reluctant to use the template for presenting their design. On the other hand, some teachers commended the clear list of important module descriptors that has to be clearly spelt out at the start of each module pedagogical pattern (the colored text in the top box in Figure 2 are placeholders, which are replaced by specific content in actual module patterns such as the patterns for Modules A and B presented to the teachers). The teachers also liked the clear specification of the learning outcomes targeted for each TLA listed in the sessions. One teacher proposed using a tabular format for listing the sequence of TLA and a separate column be created to identify the conceptual focus for each listed TLA. One teacher also suggested that a table containing all the concepts and generic skills listed in the C&A Guide should be made readily available for consultation by teachers when they work in iLAP to construct the module course-rooms.

Discussion

The work we have reported here is just a preliminary exploration of whether and in what ways a pedagogical pattern template may support deeper levels of sharing and collaboration on pedagogical design among teachers. It is clear from the teachers' responses that the preliminary pattern template for issue-enquiry modules we have constructed is far from perfect. On the other hand, the depth and richness reached by discourse is impressive, considering the very preliminary nature of this exploration. This work reveals the need for learning design support environments to scaffold teacher learning and collaboration if they are to take on more of the role of a learning design professional and less as an instructor.

Learning Design as a High-level Professional Activity

Learning design is a complex professional activity, which is multi-level and multi-faceted, as represented in Figure 3. There are design considerations to be deliberated at many levels—from entire curriculum, module, session, activity, to learning resource and technological tool to be used (if applicable). The design at each level need to take account of a number of facets: the learning goals to be achieved, the pedagogical approach to be adopted, nature of the student engagement, the kind of student-generated content to be elicited, the criteria for assessment (and rubrics to be used if applicable), and the kind of feedback that needs to be provided to students. These different facets need to be well orchestrated to achieve the design requirements at each level. These different levels of design need to be consistent and aligned with each other. The teaching profession needs appropriate learning design support environments to rise to the challenge of collaborating to improve pedagogical design.

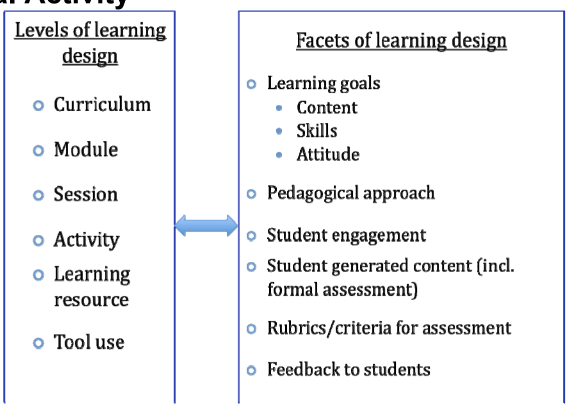


Figure 3. The Levels and Facets of Pedagogical Design.

A recent contribution to the Foresight Report points out that teaching and learning is so complex an enterprise that it needs the seamless integration of all the current forms of interactive, adaptive, informational, virtual, communication, collaborative, and presentational forms of technology currently in use (Laurillard, Kolokitha, Mellar, Selwyn, and Noss 2009). While other design-focused professionals such as architects and engineers have well developed conventions and tools for describing, constructing, sharing and communicating designs, the same is not available for teaching professionals in relation to pedagogical design. This lack of a commonly adopted convention and tools for learning design poses serious obstacles to teacher learning and collaboration, and consequently the speed with which pedagogical advances can be propagated.

Adapting, Testing and Improving Learning Designs

If teachers are to articulate their developing knowledge of effective pedagogy, then they need to be not only able to share and build on each other's work, but also to test it. This happens in the classroom every day, but if there is a viable way of setting a design against theoretical principles for good design, then to some extent its quality can be estimated in advance. A well-developed patterns template should contain all the information needed to do that, because it tells us how learners are spending their time, on what kinds of activities, and how the balance of their time is distributed across the different kinds of learning. The LDSE has an internal representation of the pedagogic type of each activity in the pattern and, using the information about the respective times spent on each activity, can interpret the nature of the whole pedagogical pattern as a pie-chart showing the distribution of the kinds of learning it affords, whether through acquisition, inquiry, discussion, practice, or production, and the distribution of learning outcomes these activities target. Building on the outcomes of the present study and the patterns and design support tools that the LDSE project has developed, we hope to construct an issue-enquiry pedagogical patterns support system that Liberal Studies teachers in Hong Kong can use as a kind of microworld for learning design, to enable teachers to design-test-redesign before trying it on their students. And because it also captures their design, they and their students can annotate it retrospectively, improve it, and then publish and share it with their peers. In this way we hope to *appropriate technology to scaffold professional collaboration and sharing in learning design.* A learning design support environment of this kind would offer a specialized tool for teachers of a kind they have never had, that exploits the capabilities of the technology to represent their decisions, interpret them and visualise the feedback. By representing the full complexity of what teachers do, in a way that remains close to their practice, and yet also challenges it, they become more empowered to develop that practice, in collaboration with their peers.

Endnotes

(1) For a discussion of issues related to low adoption, see http://opencontent.org/blog/
(2) Details can be found at the project website http://learn20.cite.hku.hk
(3) Moodle is an open source Learning Management System (LMS). Details can be found from http://moodle.org/.
(4) iLAP stands for interactive Learning and Assessment Platform. Details can be found from http://learn20.cite.hku.hk/page.php?page=platform
(5) See the project website at http://www.ldse.org.uk for more information.
(6) Learning Designs http://www.learningdesigns.uow.edu.au/
(7) ICOPER http://www.icoper.org/
(8) Available for trial at tinyurl.com/ldsepatterns

References

Biggs, J. (2003). *Teaching for Quality Learning at University*. Buckingham: SRHE/OUP.

CDC & HKEAA (2007) *Liberal Studies Curriculum and Assessment Guide.* Hong Kong: Education Bureau.

CERI. (2001). *What Schools for the Future*. Paris, France: OECD.

Goodyear, P. (2005). Educational design and networked learning: Patterns, pattern languages and design practice. *Australasian Journal of Educational Technology, 21*(1), 82-101.

Kozma, R. (2008). *ICT, Education Reform, and Economic Growth: A Conceptual Framework.* IntelCorporation.

Laurillard, D., Kolokitha, M., Mellar, H., Selwyn, N., & Noss, R. (2009). *Learning through life: The role of technology*. London: Office of Science and Innovationo. Document Number)

Laurillard, D., & Ljubojevic, D. (in press). Evaluating learning designs through the formal representation of pedagogical patterns. In J. W. Kohls & C. Kohls (Eds.), *Investigations of E-Learning Patterns: Context Factors, Problems and Solutions*: IGI Global.

Laurillard, D. & McAndrew, P. (2003). Reusable educational software: a basis for generic learning activities. In A. Littlejohn (Ed.), *Reusing Online Resources: A Sustainable Approach to e-Learning*. London: Kogan Page.

Partnership for 21st Century Skills, (2003). Learning for the 21st century: A report and MILE guide for 21st century skills. Retrieved 8/11/2010,from http://www.21stcenturyskills.org/downloads/P21_Report.pdf

Pelgrum, W. J., & Law, N. (2003). *ICT in education around the world: trends, problems and prospects*. Paris: UNESCO: International Institute for Educational Planning.

Plomp,T., Anderson,R.E., Law, N., & Quale, A. (Eds.). (2009). *Cross-national Information and Communication Technology Policy and Practices in Education* (2nd ed). Greenwich, CT: Information Age Publishing Inc.

UNESCO, 2008 *Towards information literacy indicators*. Paris: UNESCO

Acknowledgments

The Learning 2.0 project in Hong Kong is funded by the Quality Education Fund. The UK LDSE project is funded by the ESRC/EPSRC Teaching and Learning Research Programme in Technology Enhanced Learning (RES-139-25-0406). The award of a HKU Overseas Fellowship made it possible for building the research collaboration between the two institutions.

Seeing What We Mean: Co-experiencing a Shared Virtual World

Gerry Stahl, iSchool @ Drexel University, Philadelphia, US, gerry@GerryStahl.net
Nan Zhou, iSchool @ Drexel University, Philadelphia, US, nan.zhou@ischool.drexel.edu
Murat Perit Cakir, Informatics Institute, METU, Ankara, TR, perit@ii.metu.edu.tr
Johann W. Sarmiento-Klapper, SAP Research Labs, Philadelphia, US, jsarmi@gmail.com

Abstract: The ability of people to understand each other and to work together face-to-face is grounded in their sharing of our meaningful natural and cultural world. CSCL groups—such as virtual math teams—have to co-construct their shared world with extra effort. A case study of building shared understanding online illustrates these aspects: Asking each other questions is one common way of aligning perceptions. Literally looking at the same aspect of something as someone else helps us to see what each other means. The co-constructed shared world has social and temporal as well as objective dimensions. This virtual world grounds communicative, interpersonal, and task-related activities for online groups, making possible group cognition that exceeds the limits of the individual cognition of the group members.

The Shared World of Meaning

We all find others and ourselves within one world. We learn about and experience the many dimensions of this world together, as we mature as social beings. Infants learn to navigate physical nature in the arms of caregivers, toddlers acquire their mother tongue by speaking with others, adolescents are socialized into their cultures, and adults master the artifacts of the built environment designed by others. The world is rich with socially endowed meaning, and we perceive and experience it as immediately meaningful. Because we share the meaningful world, we can understand each other and can work together on concerns in common. Our activities around our common concerns provide a shared structuring of our world in terms of implicit goals, interpersonal relations, and temporal dimensions. These structural elements are reflected in our language: in references to artifacts, in social positioning, and in use of tenses. All of this is understood the same by us unproblematically based on our lived experience of the shared world. Of course there are occasional misunderstandings, particularly across community boundaries, but these are exceptions that prove the rule of shared understanding in general.

The "problem" of establishing intersubjectivity is a pseudo-problem in most cases. Human existence is fundamentally intersubjective from the start. We understand the world as a shared world and we even understand ourselves through the eyes of others and in comparison with others (Mead, 1934/1962). Rationalist philosophy—from Descartes to cognitive science—has made this into a problem by focusing on the mind of the individual as if it were isolated from the world and from other people. That raises the pseudo-problem of epistemology: how can the individual mind know about states of the world and about states of other minds? Rationalist philosophy culminated in an information-processing view of human cognition, modeled on computer architecture: understanding (as described by Dreyfus, 1992) is viewed as primarily consisting of a collection of mental representations (or propositions) of facts stored in a searchable memory.

Critiques of the rationalist approach (e.g., Dreyfus, 1992; Schön, 1983; Suchman, 1987; Winograd & Flores, 1986) have adopted a phenomenological (Heidegger, 1927; Husserl, 1936; Merleau-Ponty, 1945), hermeneutical (Gadamer, 1960/1988), or ethnomethodological (Garfinkel, 1967) approach, in which understanding is grounded in being-in-the-world-together, in cultural-historical traditions, and in tacit social practices. This led to post-cognitive theories, with a focus on artifacts, communities-of-practice, situated cognition, distributed cognition, group cognition, activity, and mediations by actor-networks. Human cognition is recognized to be a social product (Hegel, Marx, Vygotsky) of interaction among people, over time, within a shared world. Knowledge is no longer viewed as primarily mental representations of individuals, but includes tacit procedural knowledge (Polanyi, 1966), designed artifacts (Hutchins, 1996), physical representations (Latour, 1992), small-group processes (Stahl, 2006), embodied habits (Bourdieu, 1972/1995), linguistic meanings (Foucault, 2002), activity structures (Engeström, Miettinen & Punamäki, 1999), community practices (Lave, 1991), and social institutions (Giddens, 1984). The critique of human thought as purely mental and individual is now well established for embodied reality. But what happens in virtual worlds. Where the physical world no longer grounds action and reflection? That is the question for this paper.

Constructing a Shared Virtual World

However, the problem of shared understanding rises again—and this time legitimately—within the context of computer-supported collaborative learning (CSCL). That is because when students gather in a CSCL online environment, they enter a virtual world, which is distinct from the world of physical co-presence. They leave the world of nature, of physical embodiment, of face-to-face perception. They enter a world that they have not all grown into together. But this does not mean that "shared understanding" is just a matter of overlapping opinions of mental models for online groups either.

In the Virtual Math Teams (VMT) Project, we have been studying how students interact in a particular CSCL environment designed to support online discourse about mathematics. In this paper we will illustrate some of our findings about how interaction in the VMT environment addresses the challenge of constructing a shared virtual world, in which small groups of students can productively engage in collaborative mathematics.

This paper will present a case study of Session 3 of Team C in the VMT Spring Fest 2006. Here, students aged 12-15 from different schools in the US met online for four hour-long sessions. Neither the students nor the researchers knew anything about the students other than their login user names and their behavior in the sessions. A researcher joined the students, but did not engage with them in the mathematics. Between sessions, the researchers posted feedback in the shared whiteboard of the environment. The VMT Project is described and discussed in (Stahl, 2009); its theoretical motivation is presented in (Stahl, 2006). The VMT environment is shown in Figure 1. The complete chat log of Session 3 of Team C is given in the Appendix of the online version of this paper (http://GerryStahl.net/pub/cscl2011.pdf) and a Replayer version can be obtained from the authors.

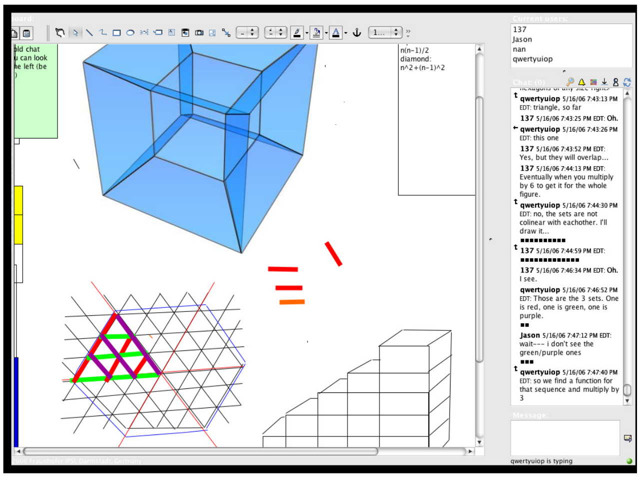


Figure 1. The VMT Environment during Session 3.

In the next sections, we illustrate the following aspects of building shared understanding: (a) Asking each other questions is one common way of resolving or avoiding troubles of understanding and aligning perceptions. (b) Literally looking at the same aspect of something as someone else helps us to see what each other means. (c) The co-constructed shared world has social and temporal as well as objective dimensions. (d) This world grounds communicative, interpersonal, and task-related activities for online groups.

Questioning to Share Understanding

We have analyzed how questions posed in the VMT environment often work to initiate interactions that resolve troubles of understanding and deepen shared understanding (Zhou, 2009; 2010; Zhou, Zemel & Stahl, 2008). This is in contrast to the rationalist assumption that questions are requests for propositional information. We will here review a number of questions from Session 3 of Group C and indicate how they lead to shared understanding. Unfortunately, due to space limitations, we will not be able to provide the full context for these questions or a detailed conversation analysis.

The question by Qwertyuiop in Log 1 serves a coordination function, making sure that all the students have read the feedback to Session 2 before any work begins in the new Session. This is an effort, taking the form of a question, to maintain a shared experience by having everyone take this first step together.

Log 2 is part of a complicated and subtle process of co-constructing shared understanding. It is analyzed in detail in (Çakır, Zemel & Stahl, 2009). The student named 137 has attempted to construct a grid of triangles in the whiteboard (similar to those in the lower left corner of Figure 1). He (or she) has failed (as expressed by the ironic "Great"), and has erased the attempt and solicited help by posing a question. Qwertyuiop requests clarification with another question and then proceeds to draw a grid of triangles by locating and then tweaking three series of parallel lines, following much the same procedures as 137 did. Qwertyuiop's understanding of 137's request is based not only on the "Yeah…" response to his/her "just a grid?" question, but also the detailed

Log 1.

Chat Index	Time of Posting	Author	Content
685	19:06:34	qwertyuiop	has everyone read the green text box?
686	19:06:44	Jason	one sec
687	19:06:45	137	Yes...
688	19:07:01	Jason	alright im done

Log 2.

694	19:11:16	137	Great. Can anyone m ake a diagram of a bunch of triangles?
695	19:11:51	qwertyuiop	just a grid?
696	19:12:07	137	Yeah...
697	19:12:17	qwertyuiop	ok...

Log 3.

698	19:14:09	nan	so what's up now? does everyone know what other people are doing?
699	19:14:25	137	Yes?
700	19:14:25	qwertyuiop	no-just making triangles
701	19:14:33	137	I think...
702	19:14:34	Jason	yeah
703	19:14:46	nan	good:-)
704	19:14:51	qwertyuiop	triangles are done

sequentially unfolding visual presentation of 137's failed drawing attempt.

In Log 3, the moderator, Nan, asks a question to make visible in the chat what members of the group are doing. Qwertyuiop is busy constructing the requested grid in the whiteboard and the others are presumably watching that drawing activity and waiting for its conclusion. The students do not seem to feel that there is a problem in their understanding of each other's activities. However, due to the nature of the virtual environment—in which the attentiveness of participants is only visible through their chat and drawing actions—Nan cannot know if everyone is engaged during this period of chat inaction. Her question makes visible to her and to the students the fact that everyone is still engaged. The questioning may come as a minor interference in their group interaction, since Nan's questioning positions her as someone outside the group ("everyone"), exerting authority by asking for an accounting, although it is intended to increase group shared understanding ("everyone know what other people are doing").

See What I Mean

Studies of the use of interactive whiteboards in face-to-face classrooms have shown that they can open up a "shared dynamic dialogical space" (Kershner et al., 2010) as a focal point for collective reasoning and co-construction of knowledge. Similarly, in architectural design studios, presentation technologies mediate shared ways of seeing from different perspectives (Lymer, Ivarsson & Lindwall, 2009) in order to establish shared understanding among design students, their peers, and their critics. Clearly, a physical whiteboard that people can gather around and gesture toward while discussing and interpreting visual and symbolic representations is different from a virtual shared whiteboard in an environment like VMT.

We have analyzed in some detail the intimate coordination of visual, narrative and symbolic activity involving the shared whiteboard in VMT sessions (Çakir, 2009; Çakır, Stahl & Zemel, 2010; Çakır, Zemel & Stahl, 2009). Here, we want to bring out the importance of literally looking at some mathematical object together in order to share the visual experience and to relate to—intend or "be at"—the object together. People often use the expression "I do not see what you mean" in the metaphorical sense of not understanding what someone else is saying. In this case study, we often encounter the expression used literally for not being able to visually perceive a graphical object, at least not being able to see it in the way that the speaker apparently sees it.

While empiricist philosophy refers to people taking in uninterpreted sense data much like arrays of computer pixels, post-cognitive philosophy emphasizes the phenomenon of "seeing as." Wittgenstein notes that one sees a wire-frame drawing of a cube not as a set of lines, but as a cube oriented either one way or another (Wittgenstein, 1953, sec. 177). For Heidegger, seeing things as already meaningful is not the result of cognitive interpretation, but the precondition of being able to explicate that meaning further in interpretation (Heidegger, 1927/1996, pp. 139f). For collaborative interpretation and mathematical deduction, it is clearly important that the participants see the visual mathematical objects as the same, in the same way. This seems to be an issue repeatedly in the online session we are analyzing as well.

137 proposes a mathematical task for the group in line 705 of Log 4. This is the first time that the term, "hexagonal array," has been used. Coined in this posting, the term will become a mathematical object for the group as the discourse continues. However, at this point, it is problematic for both Qwertyuiop and Jason. In line 706,

Log 4.

705	19:15:08	137	So do you want to first calculate the number of triangles in a hexagonal array?
706	19:15:45	qwertyuiop	What's the shape of the array? a hexagon?
707	19:16:02	137	Ya.
708	19:16:15	qwertyuiop	ok...
709	19:16:41	Jason	wait-- can someone highlight the hexagonal array on the diagram? i don't really see what you mean...
710	19:17:30	Jason	hmm.. okay
711	19:17:43	qwertyuiop	oops
712	19:17:44	Jason	so it has at least 6 triangles?
713	19:17:58	Jason	in this, for instance

Log 5.

714	19:18:53	137	How do you color lines?
715	19:19:06	Jason	there's a little paintbrush icon up at the top
716	19:19:12	Jason	it's the fifth one from the right
717	19:19:20	137	Thanks.
718	19:19:21	Jason	there ya go :-)
719	19:19:48	137	Er... That hexagon.

Log 6.

720	19:20:02	Jason	so... should we try to find a formula i guess
721	19:20:22	Jason	input: side length; output: # triangles
722	19:20:39	qwertyuiop	It might be easier to see it as the 6 smaller triangles.
723	19:20:48	137	Like this?
724	19:21:02	qwertyuiop	yes
725	19:21:03	Jason	yup
726	19:21:29	qwertyuiop	side length is the same...
727	19:22:06	Jason	yeah
728	19:22:13	Jason	so it'll just be x6 for # triangles in the hexagon
729	19:22:19	137	Each one has 1+3+5 triangles.

Qwertyuiop poses a question for clarification and receives an affirmative, but minimal response. Jason, unsatisfied with the response, escalates the clarification request by asking for help in seeing the diagram in the whiteboard *as* an "hexagonal array," so he can see it *as* 137 sees it. Between Jason's request in line 709 and acceptance in line 710, Qwertyuiop and 137 work together to add lines outlining a large hexagon in the triangular array. Demonstrating his ability to now see hexagons, Jason thereupon proceeds with the mathematical work, which he had halted in the beginning of line 709 in order to keep the group aligned. Jason tentatively proposes that every hexagon "has at least 6 triangles" and he makes this visible to everyone by pointing to an illustrative small hexagon from the chat posting, using the VMT graphical pointing tool.

In Log 5, 137 asks the group to share its knowledge about how to color lines in the VMT whiteboard. Jason gives instructions for 137 to visually locate the appropriate icon in the VMT interface. Demonstrating this new knowledge, 137 changes the colors of the six lines outlining the large hexagon, from black to blue, making the outline stand out visually (see Figure 1). 137 thereby finally clarifies how to look at the array of lines *as* a large hexagon, a task that is more difficult than looking at the small hexagon that Jason pointed to. In this excerpt, the group shares their working knowledge of their virtual world (the software functionality embedded in it), incidentally to carrying out their task-oriented discourse within that world.

In Log 6, Jason proposes a specific mathematical task for the group to undertake, producing a formula for the number of triangles in an hexagonal array of any given side length. (As we shall see below, the group uses the term "side length" as the measure of a geometric pattern at stage n.) Qwertyuiop responds to this proposal with the suggestion to "see" the hexagon (of any size) as a configuration of six triangular areas. (To see what Qwertyuiop is suggesting, look at Figure 1; one of the six triangular areas of the large hexagonal array has its "sticks" colored with thick lines. Looking at this one triangular area, you can see in rows successively further from the center of the hexagon a sequence of one small triangle, then three triangles, then five triangles.)

In line 723, 137 seeks confirmation that he is sharing Qwertyuiop's understanding of the suggestion. After posting, "Like this?" with a reference back to Qwertyuiop's line 722, 137 draws three red lines through the center of the large hexagon, dividing it visually into six triangular areas. Upon seeing the hexagon divided up by 137's lines, Qwertyuiop and Jason both confirm the shared understanding. Now that they are confident that they are all seeing the mathematical situation the same, namely *as* a set of six triangular sub-objects, the group can continue its mathematical work. Jason draws the consequence from Qwertyuiop's suggestion that the formula for the number of small triangles in a hexagon will simply be six times the number in one of the triangular areas of that hexagon, thereby subdividing the problem. 137 then notes that each of those triangular areas has 1+3+5 small triangles, at least for the example hexagonal array that they are looking at. The fact that the three members of the group take turns making the consecutive steps of the mathematical deduction is significant; it demonstrates that they share a common understanding of the deduction and are building their shared knowledge collaboratively.

The observation, "Each one has 1+3+5 triangles," is a key move in deducing the sought equation. Note that 137 did not simply say that each triangular area had nine small triangles. The posting used the symbolic visual representation, "1+3+5." This shows a pattern of the addition of consecutive odd numbers, starting with 1. This pattern is visible in the posting. It indicates that 137 is seeing the nine triangles *as* a pattern of consecutive odd numbers—and thereby suggests that the reader also see the nine triangles *as* such a pattern. This is largely a visual accomplishment of the human visual system. People automatically see collections of small numbers of objects as sets of that specific size (Lakoff & Núñez, 2000). For somewhat larger sets, young children readily learn to count the number of objects. The team has constructed a graphical representation in which all the members of the team can immediately see features of their mathematical object that are helpful to their mathematical task. The team is collaborating within a shared virtual world in which they have co-constructed visual, narrative, and symbolic objects in the chat and whiteboard areas. The team has achieved this shared vision by enacting practices specific to math as a profession for shaping witnessed events, such as invoking math terms and drawing each others' attention to relevant objects in the scene (Goodwin, 1994). They have learned and taught each other how to work, discuss, and perceive as a group in this shared virtual world.

Dimensions of a Virtual World

There has not been much written about the constitution of the intersubjective world as the background of shared understanding, particularly in the CSCL online context. There has not been much written about the constitution of the intersubjective world as the background of shared understanding, particularly in the CSCL online context. This is largely the result of the dominance of the cognitive perspective, which is primarily concerned with mental models and representations of the world; this rationalist view reduces the shared world to possible similarities of individual mental representations. Within the VMT Project, we have analyzed the dimensions of domain content, social interaction, and temporal sequencing in the co-construction of a virtual math team's world or joint problem space (Sarmiento & Stahl, 2008; Sarmiento-Klapper, 2009a; Sarmiento-Klapper, 2009b). In this work, we have found the following conceptualizations to be suggestive: the joint problem space (Teasley & Roschelle, 1993) and the indexical ground of reference of domain content (Hanks, 1992); the social positioning of team members in discourse (Harré & Gillet, 1999) and their self-coordination (Barron, 2000); and the temporal sequentiality of discourse (Schegloff, 1977) and the bridging of temporal discontinuities.

In previous sessions, the group has tried to derive formulae for the number of two-dimensional objects (small squares or small triangles) in a growing pattern of these objects, as well as the number of one-dimensional sides, edges or "sticks" needed to construct these objects. A major concern in counting the number of sides is the issue of "overlap." In a stair-step two-dimensional pattern (like the 2-D version of the stair-step pyramid in the lower right section of Figure 1), one cannot simply multiply the number of squares by 4 to get the number of sides because many of the sides are common to two squares. In Session 1, Team C had seen that in moving from one stage to the next of the stair-step pattern most new squares only required two new sides.

In Log 7, Qwertyuiop moves on from the derivation of the number of triangles to that of the number of sides. He "bridges" back to the group's earlier in-sight that the addition of "each polygon corresponds to [an additional] 2 sides." In bridging to past sessions, we found, it is necessary to re-situate a previous idea in the current context. In line 731, Qwertyuiop is reporting that for their hexagon formula, such situating does not work—i.e., that the current problem cannot be solved with the same method as the previous problems. The group then returns to the formula for the number of triangles and efficiently solves it by summing the sequence of consecutive odd numbers using Gauss' technique—the sum of n consecutive odd integers is n(2n/2)—which they had used in previous sessions.

Log 7.

731	19:22:29	qwertyuiop	the "each polygon corrisponds to 2 sides" thing we did last time doesn't work for triangles
732	19:23:17	137	It equals 1+3+...+(n+n-1) because of the "rows"?
733	19:24:00	qwertyuiop	yes- 1st row is 1, 2nd row is 3...
734	19:24:49	137	And there are n terms so... n(2n/2)
735	19:25:07	137	or n^2
736	19:25:17	Jason	yeah
737	19:25:21	Jason	then multiply by 6
738	19:25:31	137	To get 6n^2

In Log 8, Qwertyuiop makes a particularly complicated proposal, based on a way of viewing the sides in the large hexagon drawing. He tries to describe his view in chat, talking about sets of collinear sides. Jason does not respond to this proposal and 137 draws some lines to see if he is visualizing what Qwertyuiop has proposed, but he has not. Qwertyuiop has to spend a lot of time drawing a color-coded analysis of the sides as he sees them. He has decomposed the set of sides of one triangular area into three subsets, going in the three directions of the array's original parallel lines. He can then see that each of these subsets consists of 1+2+3 sides. There are 3 subsets in each of the 6 triangular areas. Based on this and generalizing to a growing hexagonal array, which will have sums of consecutive integers in each subset, the team can derive a formula using past techniques. At some point, they will have to subtract a small number of sides that overlap between adjacent triangular areas. Qwertyuiop has proposed a decomposition of the hexagonal array into symmetric sets, whose constituent parts are easily visible. Thus, his approach bridges back to previous group practices, which are part of the shared world of the group—see the analysis of a similar accomplishment by Group B in (Medina, Suthers & Vatrapu, 2009). The hexagonal pattern, which Team C came up with on its own, turns out to be considerably more difficult to decompose into simple patterns that the original problem given in Session 1. It strained the shared understanding of the group, requiring the use of all the major analytic tools they had co-constructed (decomposing, color-coding, visually identifying sub-patterns, summing series, eliminating overlaps, etc.).

Log 8.

742	19:25:48	qwertyuiop	an idea: Find the number of a certain set of colinear sides (there are 3 sets) and multiply the result by 3
746	19:26:36	137	As in those?
747	19:27:05	qwertyuiop	no-in one triangle. I'll draw it...
748	19:28:10	qwertyuiop	those
749	19:28:28	qwertyuiop	find those, and then multiply by 3
750	19:28:50	137	The rows?
751	19:30:01	qwertyuiop	The green lines are all colinear. There are 3 identical sets of colinear lines in that triangle. Find the number of sides in one set, then multiply by 3 for all the other sets.
752	19:30:23	137	Ah. I see.

Log 9.

804	19:48:49	nan	(we got a question for you from another team, which was posted in the lobby:
805	19:48:53	nan	Quicksilver 7:44:50 PM EDT: Hey anyone from team c, our team needs to know what n was in your equations last week
806	19:49:04	Jason	oh
807	19:49:15	137	The length of a side.
808	19:49:16	qwertyuiop	was n side length?
809	19:49:33	Jason	are you talking about the original problem with the squares
810	19:49:48	137	I think nan is.
811	19:49:58	qwertyuiop	i think it's squares and diamonds

In Log 9, the group work is interrupted by an interesting case of bridging across teams. At the end of each session, the teams had posted their findings to a wiki shared by all the participants in the VMT Spring Fest 2006. During their Session 3, Team B had looked at Team C's work on a pattern they had invented: a diamond variation on the stair-step pattern. In their wiki posting, Team C had used their term, "side length." Because members of Team B did not share Teams C's understanding of this term, they were confused by the equation and discussion that Team C posted to the wiki. Team B's question sought to establish shared understanding across the teams, to build a community-wide shared world. As it turned out, Team C had never completed work on the formula for the number of sides in a diamond pattern and Team B eventually discovered and reported the error in Team C's wiki posting, demonstrating the importance of community-wide shared understanding.

Grounding Group Cognition

CSCL is about meaning making (Stahl, Koschmann & Suthers, 2006). At its theoretical core are questions about how students collaborating online co-construct and understand meaning. In this paper, we conceptualize this issue in terms of online groups, such as virtual math teams, building a shared meaningful world in which to view and work on mathematical objects.

Log 10 illustrates a limit of shared understanding, closely related to the notion of a "zone of proximal development" (Vygotsky, 1930/1978, pp. 84-91). The original stair-step pattern consisted of one-dimensional sides and two-dimensional squares. In their Session 2, Team C had generalized this pattern into a three-dimensional pyramid consisting of cubes. Now Qwertyuiop proposes to further generalize into a mathematical fourth dimension and derive formulae for patterns of one, two, three, and four-dimensional objects. He had previously imported a representation of a four-dimensional hyper-cube (see the upper area of Figure 1) into the whiteboard for everyone to see.

Log 10.

20:12:22	qwertyuiop	what about the hypercube?
20:12:33	137	Er...
20:12:39	137	That thing confuses me.
20:13:00	137	The blue diagram, right?
20:13:13	qwertyuiop	can you imagine extending it it 4 dimensions, and a square extends into a grid?
20:13:17	qwertyuiop	yes
20:13:30	137	I didn't get that?
20:13:32	qwertyuiop	I'm having trouble doing that.
20:13:45	qwertyuiop	didn't get this?
20:13:50	137	Ya.
20:15:02	qwertyuiop	If you have a square, it extends to make a grid that fills a plane. A cube fills a space. A smaller pattern of hypercubes fills a "hyperspace".
20:15:19	137	The heck?
20:15:29	137	That's kinda confusing.
20:15:43	qwertyuiop	So, how many planes in a hyper cube latice of space n?
20:16:05	137	Er...
20:16:07	qwertyuiop	instead of "how many lines in a grid of length n"
20:16:17	qwertyuiop	does that make any sense?
20:16:30	137	No. No offense, of course.

At this point late in Session 3, Jason had left the VMT environment. Qwertyuiop was unable to guide 137 to see the drawing in the whiteboard as a four-dimensional object. Apparently, Qwertyuiop had been exposed to the mathematical idea of a fourth dimension and was eager to explore it. However, 137 had not been so exposed. They did not share the necessary background for working on Qwertyuiop's proposal. This shows that tasks for student groups, even tasks they set for themselves, need to be within a shared group zone of proximal development. The stair-step problem was in their zone—whether or not they could solve it themselves individually, they were able to solve it collectively, with enough shared understanding that they could successfully work together. Their three-dimensional pyramid turned out to be quite difficult for them to visualize in a shared way. Their diamond pattern seemed to be easy for them, although they forgot to work on some of it and posted an erroneous formula. The hexagonal array required them to develop their skills in a number of areas, but they solved it nicely. However, the hyper-cube exceeded at least 137's ability (or desire) to participate.

Rationalist philosophy reduces the complexity of social human existence to a logical, immaterial mind that thinks about things by representing them internally. It confuses the mind with the brain and conflates the two. It assumes that someone thinking about a hexagon or working on a math problem involving a hexagon must primarily be representing the hexagon in some kind of mental model. But one of the major discoveries of phenomenology (Husserl, 1936/1989) was that intentionality is always the intentionality of some object and that cognition takes place as a "being-with" that object, not as a mental act of some transcendental ego. As an example, we have seen that the members of Team C are focused on the graphical image of the hexagon in their virtual world on their computer screens. They reference this image and transform it with additional lines, colors, and pointers. They chat about this image, not about some personal mental representations. They work to get each other to see that image in the same way that they see it. This "seeing" is to be taken quite literally. Their eyes directly perceive the image. They perceive the image in a particular way (which may change and which they may have to learn to see). "Seeing" is not a metaphor to describe some kind of subjective mental process that is inaccessible to others, but a form of contact with the object in the world. We may say that shared understanding is a matter of the group members being-there-together at the graphical image in the whiteboard.

Being-there-together is a possible mode of existence of the online group. The "there" where they are is a multi-dimensional virtual world. This world was partially already there when they first logged in. It included the computer hardware and software. It included the VMT Spring Fest as an organized social institution. As they started to interact, the students fleshed out the world, building social relationships, enacting the available technology, interpreting the task instructions, and proposing steps to take together. Over time, they constructed a rich world, furnished with mathematical objects largely of their own making and supporting group practices that they had introduced individually but which they had experienced as a group.

Being-there-together in their virtual world with their shared understanding of many of this world's features, the group was able to accomplish mathematical feats that none of them could have done alone. Each individual in the group shared an understanding of their group work at least enough to make productive contributions that reflected a grasp of what the group was doing. Their group accomplishments were achieved through group processes of visualization, discourse, and deduction. They were accomplishments of group cognition, which does not refer to anything mystical, but to the achievements of group interaction. The group cognition was possible because of, and only on the basis of, the shared understanding of the common virtual world. Shared understanding is not a matter of similar mental models, but of experiencing a shared world.

Of course, there are limits to group cognition, just as there are limits to individual cognition. We saw that Team C could not understand Qwertyuiop's ideas about the fourth dimension. Without shared understanding about this, the group could not engage in discourse on that topic. Group cognition can exceed the limits of the individual cognition of the group members, but only by a certain amount. The individuals must be able to stretch their own existing understanding under the guidance of their peers, with the aid of physical representations, tools, concepts, scaffolds, and similar artifacts, whose use is within their grasp—within their zone of proximal development (Vygotsky, 1930/1978). We have seen that Team C was able to solve a complex mathematical problem that they set for themselves involving a hexagonal array by building up gradually, systematically, and in close coordination a meaningful virtual world.

An analysis of the log of the interaction in our case study has demonstrated much about the team's group cognition. Their group work proceeded by contributions from different individuals, with everyone contributing in important ways. Their questions showed that their individual cognition was initially inadequate to many steps in the work; but their questions also served to expand the shared understanding and to ensure that each member shared an understanding of each step. Because, the students demonstrated an understanding of the group work through their successive contributions, we can see not only that individual learning took place, but we can analyze the interactional processes through which it took place through detailed analysis of their chat and drawing actions.

As Vygotsky argued, not only does group cognition lead individual cognition by several years, but individual cognition itself develops originally as a spin-off of group cognition. Individuals can learn on their own, but the cognitive and practical skills that they use to do so are generally learned through interaction with others and in small groups. This is a powerful argument for the use of CSCL in education. It is incumbent upon CSCL research to further analyze the processes by which this takes place in the co-construction of shared understanding within co-experienced virtual worlds. As we have seen, participants in CSCL virtual environments co-construct worlds to ground their interactions. These virtual worlds exploit meaning-making, perceptual and referential practices learned in the physical social world.

References

Barron, B. (2000). Achieving coordination in collaborative problem-solving groups. *Journal of The Learning Sciences. 9*(4), 403-436.

Çakir, M. P. (2009). *How online small groups co-construct mathematical artifacts to do collaborative problem solving.* Unpublished Dissertation, Ph.D., College of Information Science and Technology, Drexel University. Philadelphia, PA, USA.

Çakır, M. P., Stahl, G., & Zemel, A. (2010). *Interactional achievement of shared mathematical understanding in virtual math teams.* Paper presented at the International Conference of the Learning Sciences (ICLS 2010). Chicago, IL. Web: http://GerryStahl.net/pub/icls2010cakir.pdf

Çakır, M. P., Zemel, A., & Stahl, G. (2009). The joint organization of interaction within a multimodal CSCL medium. *International Journal of Computer-Supported Collaborative Learning. 4*(2), 115-149. Web: http://GerryStahl.net/pub/ijCSCL_4_2_1.pdf Doi: http://dx.doi.org/10.1007/s11412-009-9061-0

Dreyfus, H. (1992). *What computers still can't do: A critique of artificial reason*. Cambridge, MA: MIT Press.

Engeström, Y., Miettinen, R., & Punamäki, R.-L. (Eds.). (1999). *Perspectives on activity theory*. New York, NY: Cambridge University Press.

Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.

Giddens, A. (1984). Elements of the theory of structuration. In *The constitution of society.* (pp. 1-40): U of California Press.

Goodwin, C. (1994). Professional vision. *American Anthropologist. 96*(3), 606-633.

Hanks, W. (1992). The indexical ground of deictic reference. In A. Duranti & C. Goodwin (Eds.), *Rethinking context: Language as an interactive phenomenon.* (pp. 43-76). Cambridge, UK: Cambridge University Press.
Harré, R., & Gillet, G. (1999). *The discursive mind*. London, UK: Sage.
Heidegger, M. (1927/1996). *Being and time: A translation of Sein und Zeit* (J. Stambaugh, Trans.). Albany, NY: SUNY Press.
Hutchins, E. (1996). *Cognition in the wild*. Cambridge, MA: MIT Press.
Kershner, R., Mercer, N., Warwick, P., & Staarman, J. K. (2010). Can the interactive whiteboard support young children's collaborative communication and thinking in classroom science activities? *International Journal of Computer-Supported Collaborative Learning. 5*(4)
Lakoff, G., & Núñez, R. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being*. New York City, NY: Basic Books.
Latour, B. (1992). Where are the missing masses? The sociology of a few mundane artifacts. In W. E. Bijker & J. Law (Eds.), *Shaping technology/building society.* (pp. 225-227). Cambridge, MA: MIT Press.
Lave, J. (1991). Situating learning in communities of practice. In L. Resnick, J. Levine & S. Teasley (Eds.), *Perspectives on socially shared cognition.* (pp. 63-83). Washington, DC: APA.
Lymer, G., Ivarsson, J., & Lindwall, O. (2009). Contrasting the use of tools for presentation and critique: Some cases from architectural education. *International Journal of Computer-Supported Collaborative Learning. 4*(4), 423-444. Doi: http://dx.doi.org/10.1007/s11412-009-9073-9
Medina, R., Suthers, D. D., & Vatrapu, R. (2009). Representational practices in VMT. In G. Stahl (Ed.), *Studying virtual math teams.* (ch. 10, pp. 185-205). New York, NY: Springer. Web: http://GerryStahl.net/vmt/book/10.pdf Doi: http://dx.doi.org/10.1007/978-1-4419-0228-3_10
Polanyi, M. (1966). *The tacit dimension*. Garden City, NY: Doubleday.
Sarmiento, J., & Stahl, G. (2008). *Extending the joint problem space: Time and sequence as essential features of knowledge building.* Paper presented at the International Conference of the Learning Sciences (ICLS 2008). Utrecht, Netherlands. Web: http://GerryStahl.net/pub/icls2008johann.pdf
Sarmiento-Klapper, J. W. (2009a). *Bridging mechanisms in team-based online problem solving: Continuity in building collaborative knowledge.* Unpublished Dissertation, Ph.D., College of Information Science and Technology, Drexel University. Philadelphia, PA, USA.
Sarmiento-Klapper, J. W. (2009b). The sequential co-construction of the joint problem space. In G. Stahl (Ed.), *Studying virtual math teams.* (ch. 6, pp. 83-98). New York, NY: Springer. Web: http://GerryStahl.net/vmt/book/6.pdf Doi: http://dx.doi.org/10.1007/978-1-4419-0228-3_6
Schegloff, E. A. (1977). Narrative analysis, thirty years later. *Journal of Narrative and Life History. 7*(1-4), 97-106.
Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. New York, NY: Basic Books.
Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge*. Cambridge, MA: MIT Press. 510 + viii pages. Web: http://GerryStahl.net/mit/
Stahl, G. (2009). *Studying virtual math teams*. New York, NY: Springer. 626 +xxi pages. Web: http://GerryStahl.net/vmt/book Doi: http://dx.doi.org/10.1007/978-1-4419-0228-3
Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences.* (pp. 409-426). Cambridge, UK: Cambridge University Press. Web: http://GerryStahl.net/elibrary/global
Suchman, L. (1987). *Plans and situated actions: The problem of human-machine communication*. Cambridge, UK: Cambridge University Press.
Teasley, S. D., & Roschelle, J. (1993). Constructing a joint problem space: The computer as a tool for sharing knowledge. In S. P. Lajoie & S. J. Derry (Eds.), *Computers as cognitive tools.* (pp. 229-258). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
Vygotsky, L. (1930/1978). *Mind in society*. Cambridge, MA: Harvard University Press.
Winograd, T., & Flores, F. (1986). *Understanding computers and cognition: A new foundation of design*. Reading, MA: Addison-Wesley.
Wittgenstein, L. (1953). *Philosophical investigations*. New York, NY: Macmillan.
Zhou, N. (2009). Question co-construction in VMT chats. In G. Stahl (Ed.), *Studying virtual math teams.* (ch. 8, pp. 141-159). New York, NY: Springer. Web: http://GerryStahl.net/vmt/book/8.pdf
Zhou, N. (2010). *Troubles of understanding in virtual math teams.* Unpublished Dissertation, Ph.D., College of Information Science and Technology, Drexel University. Philadelphia, PA, USA.
Zhou, N., Zemel, A., & Stahl, G. (2008). *Questioning and responding in online small groups engaged in collaborative math problem solving.* Paper presented at the International Conference of the Learning Sciences (ICLS 2008). Utrecht, Netherlands. Web: http://GerryStahl.net/pub/icls2008nan.pdf

Is CSCL the Missing Link between Education and the 21st Century Economy?

Jeremy Roschelle, Charles Patton, Yukie Toyama, SRI International, 333 Ravenswood Avenue, Menlo Park, CA 94025 USA
Email: Jeremy.Roschelle@sri.com, Charles.Patton@sri.com,Yukie.Toyama@sri.com

Abstract: To foster discussion that links CSCL to policy and practice (per the conference theme), we ask: what can CSCL research contribute towards preparing learners for the future economy? Building on the knowledge base we developed in a 2-year research project involving both learning scientists and economists, we first briefly review some of the challenges in establishing strong links from educational innovations to economic impacts. We postulate that the promise of linking CSCL to economic concerns requires shifting attention from goals of efficiency to goals that balance efficiency and innovation – yet the outcomes most researchers seek and measure still focus too much on efficiency. To further catalyze discussion, we suggest more deeply theorizing innovation and developing assessments that are sensitive to cognitive diversity. We offer a definition of innovation tasks as those that require student to adapt resources and expertise to define and solve challenging problems for the benefit of others.

Introduction

This year's conference theme is "connecting computer supported collaborative learning to policy and practice." In China, as in other countries, policy leaders broadly see education as relating to the demands of a 21st century economy. For example, the official report of the Seventeenth CPC Central Committee Plenary Meeting of the Fifth which took place in October, 2010) pointed out that the next five years is the crucial period of economic development of China. In order to promote long-term stable and rapid economic development, China's leaders believe that China needs to further deepen the implementation of the "Rejuvenating the Nation through Science and Education Strategy", encourage scientific and technological innovation, and improve the existing science and technology innovation system and mechanism to speed up educational reform and development (http://tinyurl.com/ChinaEdEconPolicy). Similar beliefs are also articulated in the U.S. National Educational Technology Plan (US Department of Education, 2010) which opens by stating "Education is the key to America's economic growth and prosperity and to our ability to compete in the global economy." A wide range of policy documents from the range of geographic regions represented at a typical CSCL conference state this policy concern (e.g. Committee on Science, Engineering, and Public Policy, 2007, OECD 2006).

We acknowledge that CSCL researchers can adopt many valid rationales for their research, not limited to economic concerns. Although researchers are prone to retell the story of educational reform in purely economic terms, a major concern driving educational reform in the United States one hundred years ago was social assimilation of a booming immigrant population. At other times, nations have pursued reforms to increase their ability to reproduce the best of their cultural accomplishments. In the civil rights era, the major goals of educational reform were to increase the equity of access to educational opportunity. Although, transitioning to a future economy need not be the only way to link CSCL to policy and practice; it does appear to be what is on policymakers' minds right now.

The challenge of connecting CSCL to policy through the lens of economic concerns offers CSCL researchers a chance to operate in "Pasteur's Quadrant" (Stokes, 1997) -- to conduct research that is simultaneously foundational and applied to compelling societal needs. Much CSCL research emphasizes rather foundational concerns with how we can analyze collaborative discourse (Bohr's Quadrant); other CSCL research inventively explores the potential of new technology without necessarily advancing theory (Edison's Quadrant). What would it take for CSCL researchers deeply operate in Pasteur's Quadrant where the compelling societal need is the pressure leaders feel to prepare learners for an emerging Innovation Economy?

We phrase this last question with deliberate caution, because we believe that it is easy to superficially link CSCL research to a 21st century economy ("in the future, we all need to use technology and collaborate more") while overlooking the very real gaps in our knowledge base and methods – for example, that almost all CSCL research consists of cross-sectional studies (how students improve from pre-test to post-test over a short time period) and we have almost no knowledge of whether CSCL-based improvements persist beyond the initial intervention and have on-going longitudinal consequences in students' lives. If we are serious about operating in Pasteur's Quadrant, we need to advance knowledge and methods to make links between collaborative learning and the economic needs stronger and more plausible.

To engage the community in discussing this challenge, this paper shares highlights of a much longer manuscript to appear in the Journal of the Learning Sciences (Roschelle, Bakia, Toyama, & Patton, forthcoming) and further develops themes of collaborative learning and innovation beyond what we wrote in that paper. The JLS paper represents the work of a team of scholars with backgrounds both in the learning sciences and the economy to make sense of the existing literature, to probe the strength of existing evidence linking education and the economy, and to seek direction for how to strengthen research in the future. We advance beyond that paper by focusing here on how we could more deeply theorize connections among CSCL, innovation, and the needs of a 21st century economy.

Reconsidering Outcomes: Innovation and Efficiency

A focus on efficiency is pervasive in education and perhaps is most easily recognized by its distinctive pattern of means-ends thinking. In an efficiency orientation, the end (the goal) of an educational reform is known and set ahead of time—for example, we want to increase a particular test score. The question is how to accomplish that end most efficiently. For example, in a paper (Roschelle et al, 2009) that won an award at the last CSCL conference, we asked "will a mobile CSCL technology increase the efficiency with which students learn the challenging topic of fractions?" In this question, CSCL is a means and the end is a test score. In the preponderance of CSCL research, the measured outcome is how efficiently (e.g. with lower standard deviation and greater mean learning gains) did students achieve a defined knowledge or skill – and collaboration is seen as a mediating process.

A distinction between means-ends thinking and innovation goes back at least as far as John Dewey, who distinguished between ordinary problem solving and inquiry (Dewey, 1938). In inquiry, the problem frame itself is open to reconfiguration – new perceptions can be registered, new relationships formed, and new ends visualized. More recently, work of the LIFE Center team has resulting in an understanding that efficiency and inquiry are NOT in opposition or different modalities, but rather they can be two dimensions of the same flow of activity (Schwartz, Bransford & Sears, 2005). In reviewing the literature on "adaptive expertise," the team found that there is an "optimal corridor" of activity that balances between innovation and efficiency (Figure 1).

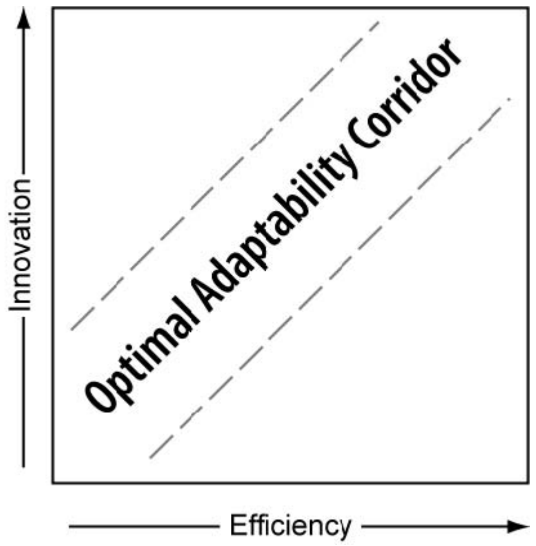


Figure 1. The Optimal Adaptability Corridor.
Source: Schwartz, Bransford, & Sears (2005).

We suggest that a concept of CSCL as a better means to efficient educational outcomes (e.g. raising today's standardized test scores) is highly unlikely to lead to lasting economic impact for participating children. However, we will also not advocate a concept of CSCL that is only focused on helping children learn to be more innovative. Instead, we suggest we need a CSCL that leads educators to be able to sustain long-term work with learners in the optimal corridor. A reformed education system cannot focus children on efficiency in routine problem solving (even through collaborative means) for a dozen years and then expect them to suddenly become innovative. Likewise, innovation requires and builds upon certain kinds of efficiency – therefore, we believe that a CSCL that offers isolated "innovation experiences" will have low net impact as well.

Now having set the stage, we broaden our perspective to consider the general problem of linking education and the economy, and then, with this in mind, consider what steps the field might have to take to move productively forward.

Difficulties in Linking Education and the Economy

In our JLS paper, which will be in print by the time of the conference, we problematize popular socio-political discourse around education and the economy. This discourse operates by coupling a "fear of falling" with a focus on education's contribution to the economy in isolation from other more powerful factors, such as the quality of management, taxation regimes, regulatory environments, etc. If education were really the only factor that produces wealth in a society, then it becomes hard to explain why Brazil has people who are educated enough to build top-rated airplanes but not to run banks. Or why an automobile in Fremont, California went from the least productive to the most productive without changing its US-educated workers (the plant changed management from GM to a joint effort of GM and Toyota). We recapitulate three key ideas from our problematization, to set a context for the rest of the paper.

Rank Is Not Likely to Be the Answer

One commonplace context in which education is linked to the economy is around international test score rankings, such as the rankings on TIMSS or PISA. A focus on rankings overemphasizes the mean differences and underemphasizes the standard deviations. Indeed, newspapers tend to report national educational rankings as if they were as straightforward as baseball team rankings, when in fact the margin of effort around these rankings is so large that all one can really say is whether a country is in the top, middle, or bottom third. Second, a focus on improve a national ranking tends to align with educational research that is purely aimed at efficiency. Increasing mean scores is emphasized and decreasing societal gaps is de-emphasized even though gaps and not means may be a more important driver of the economy, as argued by Goldin & Katz (2008). More to the point, cognitive diversity may be the most important innovation asset a society has and societies with high mean test scores (e.g. in Asian) tend to struggle with how to unleash creativity and innovation. Just as medical research has helped broad populations to understand the difference between good cholesterol and bad cholesterol, CSCL may have a role in helping societies understand that achieving high standardization of human capital (and correspondingly low standard deviations in test scores) may be counterproductive: there is a difference between bad achievement gaps and good cognitive diversity.

The Answer Is Not Necessarily in School

Our literature review also found that employers tend to value social and collaboration skills very highly, once students have achieved a sufficient educational credential (Bowles & Gintis, 2020). For example, in the Toyota management methods that improved autoworkers productivity in the Fremont, California plant, workers collaborate to improve vehicle production quality. The Toyota management did not need higher math or science scores to produce more innovative autoworkers; they needed workers who could solve problems collaboratively (Lewis, 2004). Further, we observe that as tests such as PISA introduce more real world problem solving, the test creators readily acknowledge that the tests are no longer measuring the quality of school; rather PISA is said to measure the cumulative learning opportunities a society provides to children through the age of 15. Consequently, it is not necessarily the case that the impact of CSCL that eventually reaches the economy must be mediated by in-school learning experiences, nor of CSCL applications that focus on school-like knowledge. It could be that CSCL's most important impacts will be felt outside of school. In a small example, one of the authors has observed the effects of a research-based program on youth sports in his region (http://www.positivecoach.org/); by emphasizing collaborative improvement of sports skills and teamwork rather than "winning," sports have been transformed to provide an environment in which kids learn a lot more about leveraging each others' diverse skills instead of focusing on the young star athletes.

Longitudinal Programs May Matter Most

Our literature review also noted that educational researchers tend to implicitly trust in "pipeline" concept; that students are in a pipeline from childhood to productive STEM careers and that our job is to improve one section of the pipeline. There are indeed a few studies that show that very positive early experiences can have dramatic long-term effects. For example, in the famous Perry Preschool Study (Schweinhart, Montie, Xiang, Barnett, Belfield, &. Nores, 2005), young children experienced a school setting that valued creativity and social skills as well as academic skills and parent participation. Years later, the students had higher high school graduation rates and achievement levels—and at age 27, participating students had high incomes than a comparison group (and groups were initially assigned to the Perry Preschool Program randomly). However, by and large it is rare to find longitudinal impacts from interventions that are designed only to increase a short-term measure and that are only tested cross-section ally. Indeed, most educational interventions have very poor prospects for making any noticeable difference longitudinally. Even more importantly, because we rarely measure longitudinal impacts, our field would have no way of knowing whether our best designs influence participating students' trajectories later in their studies or in their careers. We note that CSCL could have an advantage relative to subject matter specific interventions; our approaches could span multiple subject matters and multiple years. To realize this impact, we may need to focus less on improving a particular, short-term test score and more upon what it takes to achieve consistent engagement in a productive CSCL environment over the course of many years and across many topics.

Innovation and Efficiency in Existing CSCL Approaches

In this section, we discuss three existing CSCL approaches that balance goals of innovation and efficiency. Subsequently, we discuss how researchers might go beyond these approaches to more thoroughly address innovation.

Knowledge Building

Scardamalia and Bereiter's knowledge building approach (Scardamalia & Bereiter, 2006) aims at enabling school communities to attain a knowledge building character similar to that of innovation-generating organizations by providing teachers and students with an online environment called “Knowledge Forum” and a set of pedagogical principals. In this approach, all ideas are viewed as improvable and students are seen as capable of actively participating in collective idea improvement process. The online communal knowledge spaces and discourse tools make the idea-improvement process visible to everyone, thereby enabling students to contribute to collective knowledge base just as scientists and professionals of creative work do. To capture student progress in Knowledge Forum, researchers have used not only traditional summative assessment measures such as standardized tests (Scardamalia et al., 1992) but also novel formative tools such as social network analysis tools (Teplovs, Donohaue, Scardamalia, & Philip 2007; Zhang, Scardamalia, Reeve & Messina, 2009). More specifically, the latter type of measures capture the degree to which students take *collective cognitive responsibility* (Scardamalia, 2002)—being aware of others’ contributions, building on rising above or referencing others’ ideas, and participating in top-level planning decision making, and community coordination (Zhang, Scardamalia, Reeve, & Messina, 2009). These tools are not only used by the researchers but are also offered to the teacher and students to gain just in time feedback that transform on-going practice (Teplovs, Donohaue, Scardamalia, & Philip 2007).

Group Scribbles

Group Scribbles is a CSCL tool developed by SRI International in collaboration with the Learning Science Laboratory at the National Institute of Education, Singapore (Roschelle et al, 2007). Using Group Scribbles, a classroom can type or use a style to write virtual sticky notes. These sticky notes can be stored privately, shared with a team, or posted to a public display. The notes can be posted in flexible spatial configuration, and thus meaning can be established by their relative position or position relative to a background image (such as a map, diagram or template). Group Scribbles activities often focus deliberately on tapping into cognitive diversity within a classroom. For example, in one common demonstration activity (Roschelle et al, 2007), students are asked to generate fractions between 0 and 1. Then the classroom can focus on generating multiple representations of particular fractions that show the equivalence, for example, of 2/3 and 4/6. The variety of representations that students create is often wider than those shown in a textbook and comparing the meaning of 2/3 across representations is a very powerful cognitive activity. Presently a team in Singapore is scaling up Group Scribbles within local science and language arts classrooms, according to a Rapid Collaborative Knowledge Building (Ng, Looi, & Chen, 2008) approach that shares some similarities with Scardamalia and Bereiter's knowledge building approach. Research with Group Scribbles in Singapore has found that it can both improve test scores and dramatically increase the students’ innovativeness and participation of classroom discourse. For example, students in GroupScribbles classrooms contribute more often and generate more unique ideas, as well as being more likely to build on each others’ ideas.

Productive Failure

In a third example, Dr. Manu Kapur has done significant research establishing the concept of “Productive Failure” (Kapur, 2008). In an instructional sequence organized according to this concept, students are first placed in teams and given a difficult problem in the target domain of knowledge, such as “standard deviation” in statistics. For two class sessions, they collaboratively struggle to solve the problem but ultimately fail. They then receive two lectures on that present the target concept in manner that builds upon a learning progression from students’ naïve concepts to a more expert understanding. Comparison groups receive four lectures from an equally well-qualified instructor. In Dr. Kapur’s research, students in “productive failure” conditions consistently outperform students in more traditional conditions, particularly with regard to more advanced understanding and transfer of what they have learned (Kapur, 2008). One can also imagine that if collaborative “productive failure” was a more common feature of classroom instructional sequences, students would develop great comfort with one of the maxims of innovators everywhere: “Fail early and often.” Thus, although it is not instrumented by a CSCL tool, productive failure provides another example of an approach that straddles the optimal adaptivity corridor.

A Modern Concept of Innovation

The existing approaches we have cited still stay fairly close to the assumptions we critiqued earlier. Results focus on ranking (which approach produced higher mean learning gains?), the context is school, and the longitudinal impacts are rarely evaluated. Thus, we would suggest that the overall balance of the focus is still on efficiency of school learning in localized moments, and not on the broader, long-term gains in innovation capacity in a society's stock of human capital.

To move beyond an efficiency perspective in connecting the economy to education, in general, and CSCL in particular, we need to consider the challenge of the other dimension of the optimal adaptability corridor: innovation. What is required to meet the challenge of preparing our students to innovate? What are key characteristics of tasks that would require students to innovate? What is required to support students in their development of innovation capacity, that is, learn to innovate?

In a rubric our colleagues at SRI are using to examine teacher assignments and student work developed by schools, they have found this definition to be exceptionally productive:

Tasks that require students to innovate are ones that require them to:
Adapt **resources and expertise**
To **define and solve challenging problems**
To **benefit others**

The definition is inspired by research on what successful innovation looks like in the business world, and by examination of SRI's experience with innovation over its long history. It is framed in terms of what kinds of things we need to ask students to do (and support them in doing) if they are to be prepared to recognize and participate in innovation in work, in life, and in learning. Unhappily, student opportunities to take on tasks like these are quite rare, and, in the U.S. at least, those opportunities are quite inequitably distributed.

Norton Grubb, in his analysis (Grubb, 2008) of longitudinal data linking a wide range of factors with student success, makes the key theoretical distinction between simple and more complex resources The primary simple resource, money, is a poor predictor of the quality of schooling (Hanushek, 1989). However, money can be used for different purposes, and when school leaders use it and teachers to build more complex resources for instruction, those resources can be quite relevant to school quality. By complex, we mean not that such resources are hard to understand or use, but that they have multiple interlocking aspects and that they often require some amount of adaptation by local leaders and teachers – and, for innovation tasks, by students. Not simply a novel pedagogical approach, for example, but such an approach together with access to external (and possibly conflicting) expertise, mentoring, or coaching.

In his book, "The Difference" (Page, 2007), Scott E. Page lays out a powerful theoretical framework for understanding the key role of cognitive diversity in collaborative problem-solving and innovation. The components, mechanisms, and conditions of his theorem "Diversity Trumps Expertise" sheds light on all parts of the above definition of innovation, but none more so than on the issue of adapting resources and expertise. While we might suppose that in learning to adapt resources and expertise students would be learning to gain new knowledge quickly and expand the range of experts' tools they had available to them. And these, of course are good and important things. However, these are not at the heart of the matter. Embedded in both the resources and the expertise are a diversity of perspectives (more precisely, topologies) applicable to the problem space at hand. It is the iterative (and non-destructive) application of diverse perspectives – either by the individual or, as Page has it, collaboratively – that powers the triumph of diversity over expertise – for hard problems like innovation - and informs our view of adapting resources and expertise.

SRI's CEO, Curt Carlson, in his book on Innovation (Carlson & Wilmott, 2006) makes a strong case that an essential element of innovation not simply problem solving, but equally problem identification. Speaking economically, an innovation is only as significant as the problem it solves. The discipline of identifying problems, and iteratively validating, reframing, and revising them with the goal of perfecting their importance has clear value. Moreover, importance is a reasonable proxy for the "hard" or "challenging" condition of Page's theorem. In important problem that not also "hard" is not likely to be a problem for long. Both of these inform our view of the role of students self-defining the challenging problems.

No caring teacher would assign tasks at which their students would be assured to fail. And yet in tasks for which innovation is required, local failure is inevitable even while long-term success is assured. Even if students are adept at perfecting important problems, and adapting resources and expertise to leverage cognitive diversity, local "failures to advance a solution" are endemic to the process. Page's theorem assures that from *some* perspective, an advance is nearby and reachable with the available tools, not that from *every* perspective such an advance is at hand. Many perspectives will lead one in what can be seen, in retrospect, as exactly the wrong direction - but only in retrospect. Testing and validating these potential advances to a solution, then learning from and valuing them – whether they advance the solution directly or not – this is an aspect of solving challenging problems we term "productive failure." Outside of school settings, it is well understood that productive failure itself, especially when combined with problem (re-)identification, can be powerful driver of innovation (as in the classic story of the invention of Post-It® Notes.) But in school, the challenge of engaging students in solving important problems remains one of teaching and managing productive failure.

It may seem odd to see the final clause – "to benefit others" - in a rubric inspired by research on innovation in the business world, but through the writings of Doug Engelbart, and research by our colleagues at SRI on the social aspects of innovation, we have come to understand that its importance – from social, cognitive, and, frankly, practical perspectives. From the business perspective, as noted, an innovation is only as significant as the problem it solves. And that significance does not belong to the innovator but to the owner of the problem – the customer. It's not an innovation unless, and until, it "adds value" for the customer. But that's not the end of the story. In problem identification, a particularly powerful use of cognitive diversity is "seeing the problem through the eyes of the intended beneficiary." Moreover, as Page points out, while cognitive diversity trumps expertise, goal diversity (diversity in "utility" or "goodness" metrics on potential solutions) wreaks havoc on collaborative problem-solving. Externalizing the beneficiary provides a transparent – and validatable – approach to achieving commonality of "goodness" metric across all the collaborators, and enabling the development of a trust network among participants. All of these inform our view of the role of externalizing benefits of innovation.

We observe that a deeper view of innovation would distinguish a CSCL of innovation from a CSCL that purely focuses on "shared meaning" or a "community of practice" as well as from more generic accounts of 21st century skills. We suggest that further developing a CSCL of innovation would more strongly address issues of policy and practice.

Assessment: A Key Challenge on the Road Ahead

To develop a CSCL that more tightly links to innovation and the economy, we believe that CSCL researchers will have to focus on developing assessments that capture the qualities of "Human Capital" we are developing beyond academic knowledge and skills. For example, CSCL researchers need to continue developing and refining assessments to capture student capacities for collaborative problem finding and problem solving. Some of these assessments need to be summative in nature for accountability purposes, but others need to be formative, providing meaningful feedback to teachers, students, and designers of learning environments to improve their on-going practices. Additionally, these assessments need to be able to capture not only pre-determined set of student knowledge, skills, and abilities, but also their emerging capacities as they engage in CSCL. Two lines of work are particularly noteworthy to further advance this effort. First is "preparation for future learning" (PFL) approach to assessments. This approach reconceptualizes the idea of transfer from direct application of well-defined routines and facts to students' abilities to learn in a new setting with greater agility and success (Schwartz & Bransford, 1998; Schwartz, Lindgren, & Lewis, 2009). To accomplish this, learning environments must balance goals of "efficiency" and "innovation" (Schwartz, Bransford, & Sears, 2005) in order to help learners become "adaptive experts" (Hatano, 1988). Schwartz, Bransford, and Sears not only theorized about innovation and efficiency, but they also introduced a measurement paradigm for "preparation for future learning," allowing measurement of the extent to which an environment prepares students for future learning. Assessments based on the PFL approach may provide opportunities for students to try out hunches to "find" or "frame" a problem, receive feedback, and revise their work using the feedback. Such assessments may also focus on learning process and trajectories by, for example, evaluating the sophistication of questions learners ask about a topic or assumptions that they reveal in their discussions (Bransford et al., 2006). Data from such assessments are likely to provide better linkage between learning and innovation.

A second line of work that is noteworthy is the ongoing work by Knowledge Forum researchers to develop computer-assisted content analysis tools to visualize group cognition (Sha, Teplovs, & van Aalst, 2010; Teplovs & Scardamalia, 2007). This work employs Latent Semantic Analysis and Social Network Analysis to create a visual representation of student-generated notes in terms of their structural as well as semantic relationships. Uniqueness of this approach lies in its focus not only on the surface relation between student actions (i.e., note reading and writing) but also on the relation between student ideas. The approach also takes advantage of computer-automated processes to reduce labor intensity required for this type of analysis. Resulting visualizations enable one to assess how student online discussions are unfolding in terms of shared mean making and weather their ideas are becoming more coherent with one another over time. Moreover, the tools enable an examination of similarities between student discourse and the curriculum guidelines or expert discourse.

Conclusion

In this paper, we asked what it might take for CSCL researchers to operate more deeply in Pasteur's Quadrant, with foundational advances in understanding of collaborative learning and important policy and practice advances with respect to developing human capital for the evolving economy. We noted that it is easy to superficially advocate a constellation of collaboration and technology as relevant to the future, but more difficult to establish an actual causal chain from the kind of research we do to the pressing questions of policy and practice.

We have argued that for CSCL to link education to the economy, the field could productively focus on the optimal adaptivity corridor – the balance between goals of innovation and efficiency in the approaches we advocate. This focus would downplay a view of CSCL as a means only to achieving higher mean knowledge gains or to reduce standard deviations in knowledge and thus further standardize our stock of human capital. When trying to show a statistically significant effect, a large standard deviation is bad. But not all cholesterol is the same, and neither is all spread among learners to be minimized; indeed respecting and growing cognitive diversity may be one of the most important factors in increasing the innovativeness of a society. Thus a future CSCL might downplay rankings as the key measure we seek to influence and instead provide alternative measures of the cognitive diversity we are nurturing and growing, as well as the growth in collaboration, teamwork, and problem solving skills which are much desired by employers. We suggest a concurrent need to focus our efforts longitudinally and beyond school.

We have highlighted that this pathway for CSCL is grounded in established foundations through approaches like Knowledge Building, tools like Group Scribbles, and instructional sequences like Productive Failure. Further, we have argued that insights from research on innovation in the workplace could more strongly influence and direct the future of CSCL. For example, we have suggested a definition for innovative tasks that has emerged across multiple projects at SRI: innovative tasks require students to adapt resources and expertise and to define and solve challenging problems for the benefit others. We have further suggested that a key challenge for a CSCL of innovation is to advance assessments so that we can better measure students' longitudinal growth in innovation capacity.

We see the opportunity for this conference in Hong Kong to mark a turning point in the relevance of CSCL to policy and practice. As our contribution to doing so, we suggest that CSCL needs to acknowledge the strength of its past, but also move beyond a CSCL that is merely about learning gains on standardized tests, increased shared meaning among dyads and enculturation in to a disciplinary community of practice. We need to move towards a CSCL of innovation that respects cognitive diversity, emphasizes longitudinal development of human capacities, and connects more deeply to insights about how "innovation" is more than a generic 21st century skill. By doing so, we could give policy makers theory, tools, and exemplars of what it would mean to truly prepare students for their future and give practitioners insights about how to balance the needs for efficiency and innovation in their students' learning.

References

Bowles, S., & Gintis, H. (2002). Schooling in capitalist America revisited. *Sociology of Education, 75*(2), 1–18.

Brown, A. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of Learning Sciences, 2*(2), 141-178.

Bruner, J. (1999). Postscript: Some reflections on education research. In E. C. Lagemann & L. S.

Shulman (Eds.), *Issues in education research: Problems and possibilities* (pp. 399-409). San Francisco: Jossey-Bass Publishers.

Carlson, C. R. and W. W. Wilmot (2006). *Innovation: the five disciplines for creating what customers want.*

New York: Random House.

Dewey, J. (1938). *Logic: The theory of inquiry*. New York, NY, Henry Holt.

Ehrenreich, B. (1990). *Fear of falling: The inner life of the middle class.* New York: Perennial.

Goldin & Katz (2008). *The race between education and technology.* Cambridge, MA: Harvard.

Grubb, W. N. (2008). Multiple resources, multiple outcomes: Testing the "improved" school finance with NELS88. *American Educational Research Journal, 45*(1), 104-144.Hanushek, E. A. (1989). The impact of differential expenditures on school performance. *Educational Researcher, 18*(4), 45–62.

Hawkins, J., & Pea, R. D. (1987). Tools for bridging the cultures of everyday and scientific thinking. *Journal for Research in Science Teaching, 24,* 291-307.

Kapur, M. (2008). Productive failure. *Cognition and Instruction, 26(3),* 379-424.

Lave, J. (1987). *Cognition in practice*. New York: Cambridge University Press.

Lewis, W. (2004). *The power of productivity: Wealth, poverty, and the threat to global stability.* Chicago: University of Chicago Press.

Ng, F.K, Looi, C.K., & Chen, W.L. (2008). *Rapid collaborative knowledge building: Lessons learned from two primary science classrooms.* Proceedings of ICLS 2008, Utrecht, Netherlands.

Organization for Economic Co-operation and Development, OECD. (2006). *The economics of knowledge: why education is key for Europe's success.* Paris: OECD.

Page, S. (2007*). The difference: How the power of diversity creates better groups, firms, schools, and societies.*

Princeton University Press, NJ.

Roschelle, J., Tatar, D., Chaudbury, S., Dimitriadis, Y., Patton, C., & DiGiano, C. (2007). Ink, improvisation, and interactive engagement: Learning with tablets. *Computer, 40*(9), 42-48.

Roschelle, J., Rafanan, K., Estrella, G., Nussbaum, M.& Claro, S. (2009). *From handheld collaborative tool to effective classroom module: Embedding CSCL in a broader design framework.* Proceedings of the 9th international conference on CSCL.

Roschelle, J., Bakia, M., Toyama, Y., & Patton, C. (forthcoming). Eight issues for learning scientists about education and the economy. *Journal of the Learning Sciences.* Accepted for publication.

Scardamalia, M., & Bereiter, C. (2006) Knowledge building: Theory, pedagogy, and technology. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 97-117). Cambridge: Cambridge University Press.

Scardamalia, M., Bereiter, C., Brett, C., Burtis, P.J., Calhoun, C., & Smith Lea, N. (1992). Educational applications of a networked communal database. *Interactive Learning Environments, 2*(1), 45-71.

Schwartz, D. L., Bransford, J. D., & Sears, D. L. (2005). *Efficiency and innovation in transfer. Transfer of learning from a modern multidisciplinary perspective,* 1-54.

Schweinhart, L.J., Montie, J., Xiang, X., Barnett, W.S., Belfield, C.R., and M. Nores. (2005*). Lifetime effects: The High/Scope Perry Preschool study through age 40.* Ypsilanti: High/Scope Press.

Sha, L., Teplovs, C., & van Aalst (2010). *A visualization of group cognition: semantic network analysis of a CSCL community.* ICLS '10 Proceedings of the 9th International Conference of the Learning Sciences.

Stokes, D. (1997). *Pasteur's quadrant: Basic science and technological innovation.* Washington, DC: Brookings Institution Press.

Teplovs, C., Donoahue, Z., Scardamalia, M. and Philip, D. (2007). *Tools for Concurrent, Embedded, and Transformative Assessment of Knowledge Building Processes and Progress.* Demonstration presented at CSCL 2007.

U.S. Department of Education. (2010). *Transforming American education: Learning powered by technology. DRAFT National Education Technology Plan 2010.* Retrieved May 17, 2010, from http://www.ed.gov/technology/netp-2010.

Acknowledgments

We thank our colleagues who participated with us in the Learning Sciences and the Economy project, particularly Marianne Bakia. In addition, thanks to Manu Kapur for spending time with us in Singapore.

A Portrait of CSCL Methodologies

Heisawn Jeong, Hallym University, Chuncheon, Korea (South), heis@hallym.ac.kr
Cindy E. Hmelo-Silver, Rutgers University, New Brunswick, US, cindy.hmelo-silver@gse.rutgers.edu

Abstract: This study analyzed the methodologies of empirical CSCL research published in seven leading journals of the field during 2005-2008 along four dimensions of research methods, that is, research design, research setting, data, and analyses. In addition, this study examined research methodologies in relation to theoretical frameworks. Through these analyses, this study provides a detailed picture of CSCL methodologies currently practiced in the field. In addition, the study also revealed a strong influence of theoretical frameworks on all aspects of research methodology.

Introduction

Computer-Supported Collaborative Learning (CSCL) is an interdisciplinary research field concerned with how people can learn together with the help of computers (Stahl, Koschmann, & Suthers, 2006). Since its inception, researchers have brought diverse theoretical perspectives and methodological approaches to the study of CSCL. As a result, the field included an uncommon range of theoretical and methodological perspectives. Traditional information processing perspectives co-exist with socio-cultural perspectives such as activity theory and situated cognition in CSCL. The infusion of different theoretical perspectives has created a great deal of excitement and numerous efforts to bridge differences in disjointed and/or seemingly incompatible theoretical traditions (Anderson, 1997; Anderson, Reder, & Simon, 1996; Greeno, 1997; Greeno & Sande, 2007; Paavola, Lipponen, & Hakkarainen, 2004). The same is true for research methodology. In much CSCL research, the goal is no longer merely to understand the phenomena but also to transform the current practices. When working in the "blooming, buzzing confusion" of classrooms, researchers often no longer have the luxury to isolate and manipulate variables as they would in the laboratories (Brown, 1992). The environment is complex with many tightly interconnected components. Controlling variables is all but impossible in such environments. As a result, working in such settings often meant abandoning scientific rigor and theoretical relevance in the past. Borrowing from design sciences such as engineering and artificial intelligence, however, the introduction of a research strategy called "Design-based research" has allowed researchers to design changes in classroom practices while working toward the goal of developing theoretical models of learning and instruction. Design-based research became a tool not only to test and refine educational designs but also a tool to carry out formative research about learning theories (Brown, 1992; Collins, Joseph, & Bielaczyc, 2004).

Distinctions between quantitative and qualitative methodologies have long existed in many fields. Quantitative research refers to the empirical investigation of numerical properties of variables and relationships among these variables (Shadish, Cook, & Campbell, 2002). Heavily influenced by logical positivism and statistical theories, the objective of quantitative research is often to test hypothesis that are true to the population at large. It was typically associated with experimental and survey studies where numerical data are collected with active manipulation of variables. In such studies, large sample sizes are important to ensure statistical significance. Quantitative research is often contrasted with qualitative research that aims to construct an in-depth understanding of human behaviors. Qualitative research is associated with descriptive methods such as case or ethnographic studies where rich data such as videotapes, verbal transcripts, and artifacts are collected and analyzed qualitatively. Analysis typically deals with a small set of data and is aimed at uncovering meaningful patterns that may be context-specific. Statistical and numeric analyses are sparsely used. Quantitative research with its longer history has been the dominant methodology in many fields of research (Hunter & Leahey, 2008; Morrow & Brown, 1994). However, this picture is changing. Qualitative research, although initially a methodology practiced in a few fields such as anthropology and sociology, has been establishing itself as a major research methodology in many fields including education and CSCL. As qualitative methods become more popular, more attempts are made to triangulate data collection and analyses, and mixed- or multi-method research has become increasingly used (Hmelo-Silver, 2003; Johnson & Onwuegbuzie, 2004). Nonetheless, there are also cautionary voices regarding mixed-method approaches and/or methodological eclecticism on the grounds that they fail to consider theoretical backgrounds and conceptual framework for research (Maxwell, 2004; Morrow & Brown, 1994; Yanchar & Williams, 2006). What this means is that traditional descriptions of qualitative or quantitative research no longer adequately describe some of the research being carried out in the field. There is a need to develop a more sophisticated understanding of the kinds of research methodology currently practiced in CSCL investigations and ways to better align them with diverse theoretical backgrounds.

This study aimed at examining recent trends in CSCL research methodology. Our analyses examined the following dimensions of CSCL research methodology: (1) Research designs, (2) Research settings, (3) Data, and (4) Analyses. These methodological dimensions were then examined in relation to theoretical backgrounds of the investigations. We analyzed empirical CSCL research papers published in seven leading journals of the field during 2005-2008. This study is an extension of Jeong and Hmelo-Silver (2010), which reported on some of the preliminary findings for subset of papers published during 2005-2007. We added additional studies published in 2008 as well as an additional coding category.

Methods

Journal Selection

Seven journals were selected for this study: (1) *International Journal of Computer Supported Collaborative Learning* (ijCSCL) (2) *Journal of the Learning Sciences,* (3) *Learning and Instruction*, (4) *Computers and Education*, (5) *Journal of Computer Assisted Learning*, (6) *International Journal of Artificial Intelligence in Education* (ijAIinEd), and (7) *Computers in Human Behavior.* The journals were selected based on a survey of 16 CSCL community leaders (e.g., CSCL committee of ICLS and the editorial board members of ijCSCL). These are all peer-reviewed journals published by well-known publishers with international authorship and readership. Articles published in the seven journals during 2005-2008 period, that is, four years of publication and three years of publication in the case of *ijCSCL* were examined for the study.

Paper Selection

Excluding non-research articles (e.g., editorials, book reviews, or obituaries), 1,423 articles were published in the seven journals during the 2005-2008 period. We screened them to identify empirical CSCL research papers. *Empirical* research was determined based on whether the study examined primary data. Secondary data analysis, simulated results, theoretical papers, and meta-analyses were excluded. The data may have been collected as part of a larger project, but the analyses and findings had to be new. *CSCL* research was determined in terms of whether learners learned collaboratively using technological tools. Learning needed to be collaborative, but as long as parts of the learning process involve interaction (e.g., collaborative discussion after individual study), it was considered as collaborative learning. The focus was on small group peer collaboration, so that student-teacher interactions or whole class discussions were excluded unless they also included small group peer collaboration. The applied technologies needed to be specific (studies examining general IT usage were not included), but did not necessarily have to be so-called collaboration technology such as e-mails or discussion boards. Interaction with computerized agents or intelligent systems were included if they involved learning. Studies that examined social and technical issues were included if they were studied in relation to CSCL. In addition to empirical CSCL papers, we included methodological papers that addressed various methodological issues related to CSCL (e.g., introduction of new methods such as Social Network Analysis, development of specific rating schemes). Studies with special population students (e.g., ADHD, autism) were excluded.

The selection process proceeded in two stages. First, initial selection of empirical CSCL papers was conducted based on the title and abstract of the paper. At this stage, we tried to be as inclusive as possible so as not to miss any potential CSCL papers and included papers if their title and abstract suggested empirical CSCL investigations. This initial screening was verified at the coding stage so that final judgment about the eligibility was made based on a more comprehensive examination of the papers. Currently, the number of the papers in the coding pool is 315, which is roughly about 22% of the papers published in the journals during 2005-2008.

Content Analyses

In order to examine research methodologies, we examined the studies along the following dimensions: (1) Research design, (2) Study setting, (3) Data, and (4) Analysis. We then examined these features of research methodology in relation to (5) Theoretical frameworks of the study. Coding categories were developed based on a combination of inductive and deductive approaches. They were initially generated top-down (e.g., using categories drawn from the submission descriptors of the 2005 CSCL conference) and later refined through a bottom-up process of multiple coding iterations. We describe coding schemes for each category of codes below.

Research Designs

Research designs were coded as: (1) Experimental, (2) Descriptive, or (3) Design-based research methods. Experimental designs refer to studies where some interventions were applied and was further divided into (a) randomized, (b) quasi-experimental, and (c) pre-post design. Descriptive studies referred to studies that aimed at describing a phenomena or case. These included case studies, surveys, and ethnographic investigations. Design-based method referred to the research strategy in which CSCL designs and interventions were investigated in theoretically-driven ways and refined progressively over several iterations. In order to be coded as design-based method, the study not only needed to design CSCL systems or environments, but also the design needed to be theoretically grounded, instantiated in specific contexts, studied and refined iteratively as part of a bigger design-based research (Barab & Squire, 2004; Brown, 1992; Collins, Joseph, & Bielaczyc, 2004). Note that design-based research refers to an approach or framework of research that can transcend the design of individual iterations that may have been either experimental or descriptive. Once a study was coded as design-based method, the design of individual iterations was not coded separately.

Research Settings

Research settings were defined as the contexts in which the research was conducted and was coded as: (1) Laboratory, (2) Classroom, or (3) Other settings. Laboratory referred to lab-like controlled settings where data collection was carried out outside the context of classrooms or other authentic learning situations. Classroom setting referred to more or less formal learning situations that were guided by teachers both within and outside of physical classroom (e.g., field trips). Other settings meant settings outside laboratories or classrooms such as workplace or informal learning environments (e.g., teacher workshop or professional conference).

Data

Data were coded as: (1) Process, (2) Outcome, and (3) Miscellaneous. Process data were further divided into (a) text-asynchronous, (b) text-synchronous, (c) video/audio, (d) log data, and (e) other. Outcome data referred to data collected to get static snapshots at the learners' cognitive and other states and could be collected at the beginning of a study (e.g., pre-test) as well as at the end. Sub-categories were: (a) multiple-choice questions, (b) open-ended questions, (c) artifacts (e.g., contents of multi-media whiteboard), and (d) other (e.g., expert ratings, final course grades). Miscellaneous data included (a) self-reports/questionnaires (e.g., demographic information, affective measures, perceived acceptance, etc.), (b) interview or focus groups, (c) field notes or observations, and (d) other (e.g., IQ, learning styles).

Analysis Methods

Analysis methods refer to the kinds of analyses carried out on the data and consisted of two general categories: (1) Quantitative and (2) Qualitative. Quantitative analyses were further coded as: (a) simple descriptive (e.g., frequencies or means, quantitative analysis of qualitative data such as coding numbers of words in messages or scoring an open-ended answers), (b) code and count, (c) inferential statistics (e.g., t-test, analysis of variance, regression), (d) modeling (e.g., log-linear analysis, structural equation modeling), (e) Social Network Analysis (SNA), and (f) other miscellaneous quantitative analysis (e.g., comparison with simulated data). Among simple descriptive, inferential, and modeling statistics, we coded the more sophisticated form of analysis since the use of inferential statics and modeling assumes descriptive statistics and inferential statistics, respectively. If a study did a code and count and then ran inferential statistics, they were coded separately, however. Qualitative analyses were further coded as: (a) (qualitative) content analysis, (b) Conversational Analysis (CA) and Discourse Analysis (DA), (c) grounded theory, (d) Interaction Analysis (IA), (e) other (e.g. narrative analysis or phenomenography), and (f) loosely defined where data was qualitatively described with some examples.

Theoretical Frameworks

Theoretical frameworks were coded as: (1) Information processing, (2) Socio-cognitive, (3) Constructivism, (4) Socio-cultural (e.g., distributed cognition, activity theory), (5) Communication, (6) Social psychology, (7) Motivation, (8) Atheoretical, and (9) Other. Information processing theory refers to traditional cognitive theories with a strong emphasis on individualistic cognitive processes such as encoding and retrieval from memory (Klahr & Kotovsky, 1989; Newell & Simon, 1972). Socio-cognitive theory refer to theories related to Piagetian notion of cognitive conflicts and conceptual change (De Lisi & Golbeck, 1999; Doise, Mugny, & Perret-Clermont, 1975). Constructivism refers to a broad range of theoretical approaches that emphasize active learner processing and knowledge construction either in individualistic and collaborative settings (Chi, 2009; von Glaserfeld, 1987). Socio-cultural theory refers a diverse range of theories such as generic Vygotskian approach, distributed cognition, or activity theory (Engeström, 2001; Hutchins, 1995). Communication theory refers to theories addressing linguistic and communicative aspects of collaboration (Krauss & Fussell, 1990). Social psychology theory refers to theories that focus on social aspects of collaboration such as status difference, gender, or group dynamics (Levine & Thompson, 1996). Motivation theory refers to theories with a focus on motivational aspects of learning addressing issues such as self-regulation (Pintrich, 1999). Atheoretical framework refers to investigations that are more or less guided by practical concerns. Other theory category refers to theories that did not fit into any of the categories that we have described.

Coding was conducted based on researchers' description of studies. If researchers described their study as "experimental" and/or "interaction analysis," then we coded them as such. In a few cases where their description was controversial, we followed a more conventional definition so that "near synchronous" interaction was coded as asynchronous interaction and that an "experiment" without any control condition was coded as a descriptive study. In unclear cases in which the authors did not explicitly state the information relevant to the coding categories (e.g., no mention of whether the study was carried out in lab or in classrooms), we relied on the contextual information presented in the paper. For example, when a study did not specify data and only stated that the number of words in asynchronous notes was analyzed, it was assumed that asynchronous text messages were collected as data (e.g., Hewitt & Brett, 2007). Multiple coding was possible depending on the features of the study. For example, when the study collected more than one type of data such as asynchronous text messages and interview data, all of them were coded. Two coders independently coded a subset (20%) of papers from the 2005-2007 set and discussed disagreements. Cohen's kappa was above .75 for all codes (Jeong & Hmelo-Silver, 2010). The coders divided the rest of the papers and are in the process of coding them independently while discussing unclear cases.

Results

Portraits of CSCL Methodologies

About 8% of the papers in the paper pool were methodology papers. Note that methodology papers may not have reported on actual data as some discussed approaches or issues in analyses (Strijbos, Martens, Prins, & Jochems, 2006; Weinberger, Stegmann, & Fischer, 2007). Because research design or settings were not always identifiable in methodology papers, the results reported in this section are based on a random sample of the papers that contained actual data, excluding methodology papers (N=265).

Research Designs

The most prevalent CSCL research design were descriptive designs (54%), followed by experimental research (37%), and design-based research (9%). Of the studies with experimental design, about half (54%) were randomized experiments, followed by quasi-experimental (31%) and pre-post designs (15%).

Research Settings

Most often, CSCL research was conducted in classrooms (73%), followed by laboratory (21%) and other settings (9%). A small proportion of studies (3%) was conducted in multiple settings, so that an experiment was first carried out in the laboratory and then in classrooms (de Jong, Kolloffel, van der Meijden, Staarman, & Janssen, 2005). Laboratory studies have been typically associated with experimental design, whereas classroom studies have been typically associated with descriptive design in the past. Although this was generally true, there were many exceptions. Of the laboratory studies, 36% were descriptive studies, whereas a similar proportion of classroom studies (31%) were experiments (Figure 1).

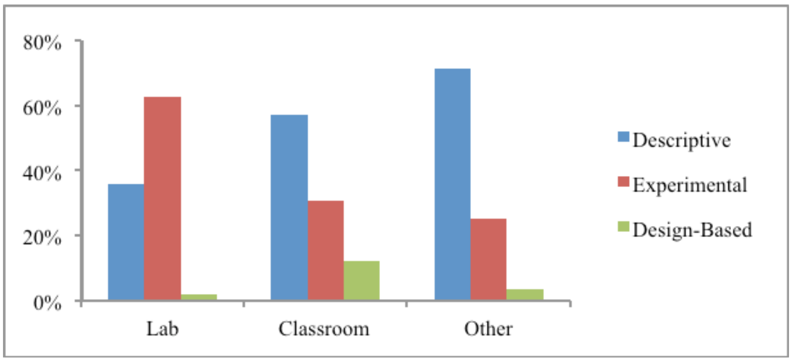


Figure 1. CSCL Research Settings by Design.

Data

Researchers collected diverse types of data in CSCL investigations. The most frequently collected process data were asynchronous text messages (29%), followed by log data (28%), video/audio (26%), synchronous text messages (15%), and other (2%). For outcomes, the most frequently collected data were artifacts (24%), followed by open-ended questions (17%), multiple-choice questions (16%), and other (15%). For miscellaneous data, most frequently collected data types were self-report (54%), interviews (30%), field notes/observations (16%), and other (8%). Overall, CSCL research collected process data in 68% of the studies, outcome data in 54% of the studies, and miscellaneous data in 72% of the studies.

Multiple data types were collected in typical CSCL investigations with the average number of data types being 2.82 per study. Interestingly, research design seemed to have influenced the amount of data collected. Design-based research collected the most types of data (*M*=3.52), followed by experimental (*M*=2.78) and descriptive (*M*=2.72) studies (Figure 2).

Analysis Methods

CSCL research used a variety of analysis methods. Overall, 86% of the studies carried out quantitative analyses and 52% of the studies carried out qualitative analyses. Percentages exceeded 100%, because 38% of the studies included both quantitative and qualitative analyses. Studies with quantitative analyses replied on descriptive statistics (20%), code and count (48%), inferential statistics (66%), modeling such as multi-level modeling (6%), SNA (1%), and other (1%). Qualitative analyses used qualitative content analysis (3%), CA or DA (9%), grounded theory (8%), and IA (2%), and other methods (10%), and loosely defined (66%).

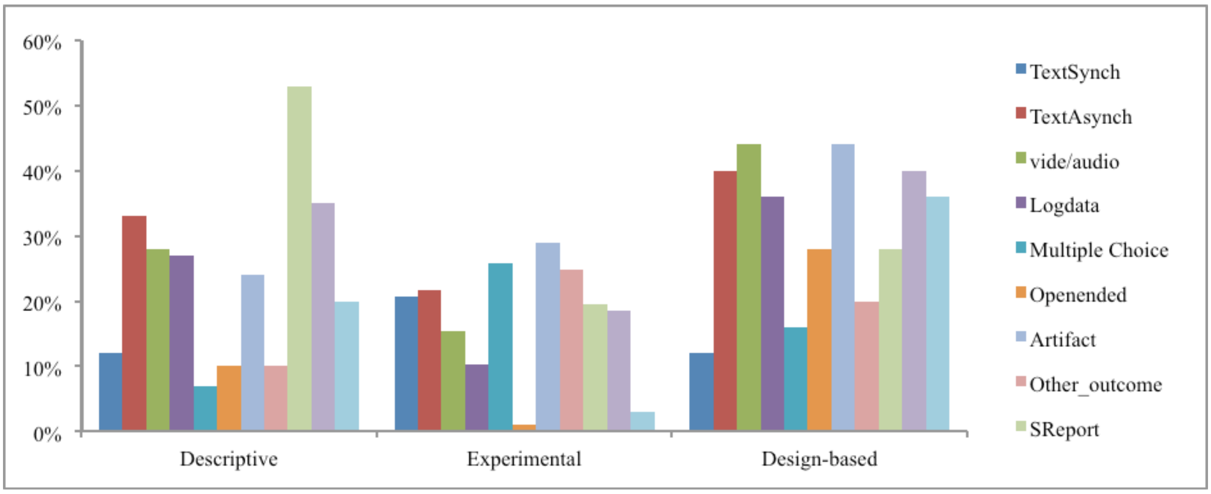


Figure 2. Data Collected in CSCL Research by Research Design (Data Types with Low Frequencies Omitted).

When researchers used quantitative analysis, they generally used inferential statistics. This was true even for studies that used code and count. The results of coding were subjected to inferential statics in 69% of the cases. At the same time, there were also a sizable amount of studies that relied on descriptive statistics only. This occurred when researchers reported some descriptive statistics in the contexts of qualitative analyses (43%) or code and count (22%), but there were many studies whose analysis was solely descriptive. In the case of qualitative analyses, the use of established techniques such as CA or DA was less common and accounted for 32% of qualitative analyses conducted. The remaining qualitative analyses were *loosely defined* where descriptions of data were provided with examples. Loosely defined analyses were usually carried out after some initial quantitative analyses were conducted. For example, qualitative, in-depth descriptions of a subset of data were provided after quantitative analyses of the whole data set. However, about 15% of qualitative analyses were solely in the loosely defined category.

Analysis methods were strongly influenced by research design (see Figure 3). While experimental and descriptive studies relied more heavily on quantitative analysis (100% and 79%, each) than qualitative analyses (28% and 57% each), design-based studies relied more heavily on qualitative analyses (84%) than on quantitative analyses (68%). Use of multiple analysis methods was most common in design-based research with about half (52%) using both quantitative and qualitative analyses. Note, however, that this could have occurred because design-based studies collected more data than other studies and, as a result, had to use a range of methods to analyze the data.

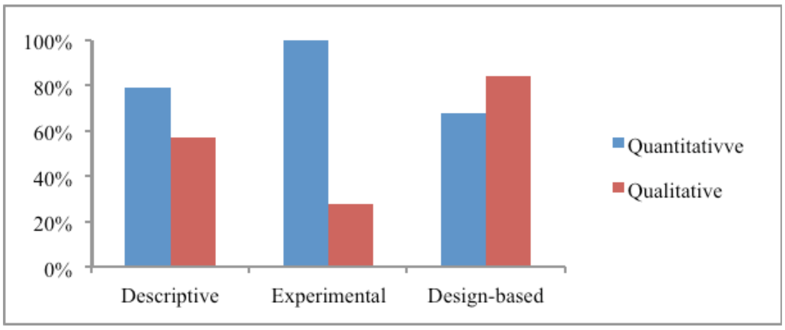


Figure 3. CSCL Research Design by Analysis Methods.

Theoretical Frameworks and Research Methodologies

CSCL research was guided by various theoretical frameworks with the most common being constructivism (32%), followed by socio-cultural theories (19%), social psychology (16%), other (14%), atheoretical (11%), information processing (10%), and socio-cognitive theories (4%). Note that there were quite a number of atheoretical investigations motivated by practical concerns. At the same time, a small portion of the papers (14%) drew from multiple theoretical frameworks, reflecting the diverse theoretical heritage of CSCL.

Theoretical frameworks had a strong influence on all aspects of the investigations. It first influenced research design (see Figure 4). Experimental studies were strongly guided by constructivism and social psychology theories, whereas descriptive and design-based studies were strongly guided by constructivism and socio-cultural theories. Theoretical frameworks similarly influenced research settings so that laboratory studies were most strongly associated with constructivism and social psychology framework, while classroom studies were strongly associated with constructivism and socio-cultural theories.

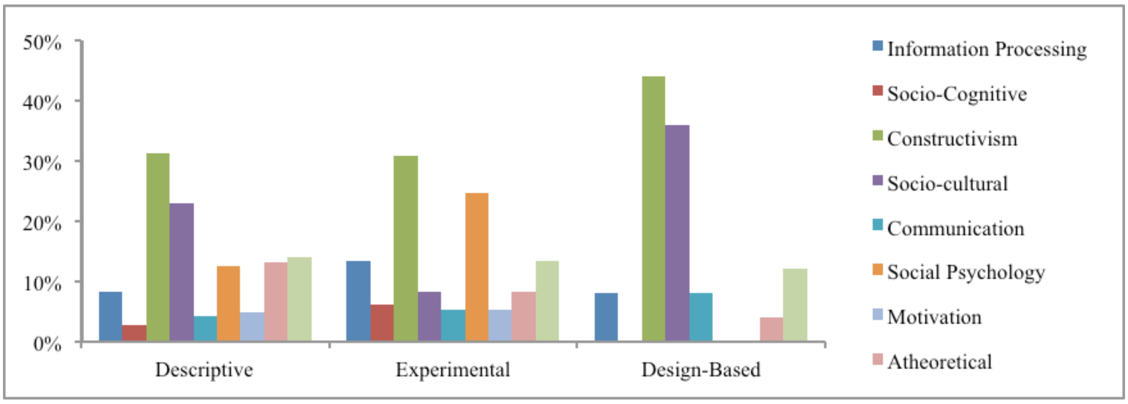


Figure 4. CSCL Research Designs by Theoretical Frameworks.

Theoretical frameworks also influenced kinds of data collected. Video/audio data and artifacts were most likely to be collected in studies with a socio-cultural framework (44% and 32% each), whereas self-report data were most likely to be collected in studies with a social psychology framework (79%). Theoretical frameworks also influenced analysis methods (Figure 5). Although quantitative analysis was less common in studies with socio-cultural frameworks (64%), quantitative analyses were actively used in studies with all theoretical frameworks. Theoretical frameworks influenced use of qualitative analysis methods more heavily. Studies from information processing and social psychology framework were least likely to carry out qualitative analyses (15% and 21% each), whereas studies from socio-cultural framework were most likely to carry out qualitative analyses (82%).

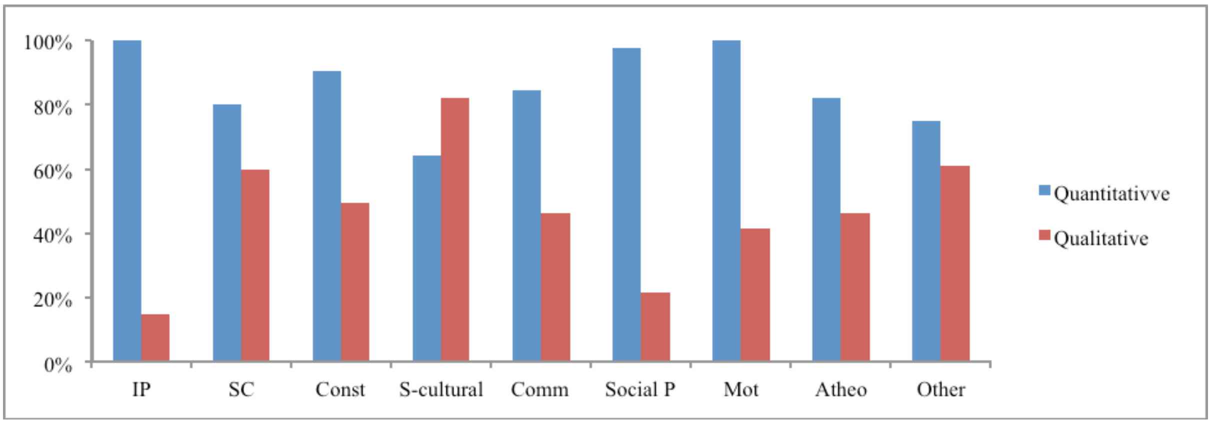


Figure 5. Theoretical Frameworks by Analysis Methods (IP: Information Processing; SC: Socio-Cognitive; Const.: Constructivism; S-cultural: Socio-cultural; Comm: Communication; Social P: Social Psychology; Mot: Motivation; Atheo: Atheoretical).

Discussion

In this paper, we conducted a detailed examination of empirical CSCL research published in seven leading journals of the field during 2005-2008. The analyses showed that the typical methodology adopted in empirical CSCL investigations is a descriptive design with quantitative analyses of self-report/questionnaire data in classroom settings. Although these may characterize the most typical CSCL empirical investigations, CSCL methodology is far from monolithic. Design-based research, although still comprising a minority of studies, has gained a respectable footing in CSCL research. Researchers are also actively experimenting with existing methodologies as they carry out experiments in classroom settings, collect many kinds of data, and increasingly analyze them both quantitatively and qualitatively. In the past, characterization of research methodology often relied on simplistic distinctions such as quantitative and qualitative and/or experimental and descriptive distinctions (Hew, Kale, & Kim, 2007; Hrastinski & Keller, 2007). Rather than relying such a general characterization, this study examined research methodology in more detail along four dimensions of research methodology: design, setting, data, and analyses and identified some of the recent trends in CSCL research. Through these analyses, this study was able to provide a more detailed picture of CSCL methodologies currently practiced in the field. Reflecting its multi-disciplinary mission, CSCL is characterized by multiple methodological approaches. The complexity of the research methodology currently practiced in CSCL suggests a need to develop more sophisticated understanding of research methodologies beyond the exiting dichotomies of quantitative and qualitative methods. In addition, this study showed that theoretical framework strongly influenced all aspects of research methodologies, suggesting that researchers need to be more mindful of the influences and biases of one's theoretical frameworks when they plan and carry out research.

Despite the recent methodological developments in the field, there are a number of areas where more sophisticated methodological understanding and practices are needed. First, the field needs to be more principled in its applications of design-based research. Researchers often did not distinguish design as a research goal from design as a research methodology. In addition, even when design research was used to refer to the research method or strategies, its application was often name in only (Jeong & Hmelo-Silver, 2010). Second, unprecedented amounts of "content" data are being generated online in the form of synchronous and asynchronous messages and various online artifacts. The advancement of the field greatly depends on how well these rich sources of data can be utilized, and much of the field's current analytic efforts is directed toward honing analytic methodologies related to analyzing these types of data, as can be seen in the topics of recent methodology papers (Baker, Andriessen, Lund, van Amelsvoort, & Quignard, 2007; de Wever, Schellens, Valcke, & van Keer, 2006). So far, analyses of these content data have been mostly conducted through coding and counting or qualitative analyses. Such methods are time-consuming and not well suited to large-scale analyses, however. Automating the process of code and count is being explored (Rosé, 2008), but the field needs to explore additional ways to analyze these data efficiently as well as meaningfully. The yearning for meaning research outcomes that goes beyond "merely significant" results is reflected in the increasing role of qualitative methodology. However, qualitative analyses often present barriers for many researchers who are new to them. It is often unclear which qualitative methods to choose among the diverse qualitative research methods such as discourse analysis and conversational analysis. The analytic techniques are not always transparent either. These factors contribute to shallow applications of qualitative methodology. Qualitative research methods need to be better communicated to the community in order to ensure more fluent and appropriate practices. Third, the mixed-method approach is becoming increasingly prevalent in CSCL research. There are concerns that the epistemological commitments of some of these methodologies are too disparate for true multi-method analyses to be possible. Epistemological issues are not necessarily present in all forms of mixed-method research. Still, researchers need to be aware of the differences in epistemological and theoretical frameworks and find ways to achieve productive multi-vocality (Suthers, et al., 2011). Fourth, in this paper, we attempted to contextualize research methodology within theoretical frameworks and captured some of its influences. In a similar fashion, we need to understand what may be the influence of research methodologies (as well as theoretical framework) on research findings. We need to understand whether and how adoption of certain methodology obscures certain aspects of the phenomena while increasing sensitivity to others. The results of this study should help researchers enhance their understanding of current practices of CSCL research methodologies as well as develop common conceptual frameworks for discussing different methodologies and bridging diverse research traditions.

References

Anderson, J. R. (1997). Situative versus cognitive perspectives: Form versus substance. *Educational Researcher, 26*(1), 18-21.

Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). Situated learning and education. *Educational Researcher, 25*(4), 5-11.

Baker, M., Andriessen, J., Lund, K., van Amelsvoort, M., & Quignard, M. (2007). Rainbow: A framework for analyzing computer-mediated pedagogical debates. *International Journal of Computer-Supported Collaborative Learning, 2*, 315-357.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of The Learning Science, 2*(2), 141-178.

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science, 1*, 73-105.

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: theoretical and methodological issues. *Journal of the Learning Sciences, 13*(1), 15-42.

de Jong, F., Kolloffel, B., van der Meijden, H., Staarman, J. K., & Janssen, J. (2005). Regulative processes in individual, 3D, and computer supported cooperative learning contexts. *Computers in Human Behavior, 21*, 645-670.

De Lisi, R., & Golbeck, S. L. (1999). Implications of Piagetian theory for peer learning. In A. M. O'Donnell & A. King (Eds.), *Cognitive perspectives on peer learning*. Mahwah: Erlbaum.

de Wever, B., Schellens, T., Valcke, M., & van Keer, H. (2006). Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. *Computers and Education, 46*, 6-28.
Doise, W., Mugny, G., & Perret-Clermont, A. (1975). Social interaction and the development of cognitive operations. *European Journal of Social Psychology, 5*(3), 367-383.
Engeström, R. (2001). Expansive learning at work: toward an activity theoretical reconceptualization. *Journal of Education and Work, 14*(1), 133-156.
Greeno, J. G. (1997). Response: On claims that answers the wrong questions. *Educational researcher, January/February*, 5-17.
Greeno, J. G., & Sande, C. (2007). Perspectival understanding of conceptions and conceptual growth in interaction. *Eduational Psychologist, 42*(1), 9-23.
Hew, K. F., Kale, U., & Kim, N. (2007). Past research in instructional technology: Results of a content analysis of empirical studies published in three prominent instructional technology journals from the year 2000 through 2004. *Journal of Educational Computing Research, 36*(3), 269-300.
Hmelo-Silver, C. E. (2003). Analyzing collaborative knowledge construction: Multiple methods for integrated understanding. *Computers and Education, 41*, 397-420.
Hrastinski, S., & Keller, C. (2007). An examination of research approaches that underlie research on educational technology: A review from 2000 to 2004. *Journal of Educational Computing Research, 36*(2), 175-190.
Hunter, L., & Leahey, E. (2008). Collaborative Research in Sociology: Trends and Contributing Factors. *American Sociologist., 39*(4), 290-306.
Hutchins, E. (1995). *Cognition in the wild.* Cambridge, MA: The MIT Press.
Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed method research: A research paradigm whose time has come. *Educational Researcher, 33*(7), 14-26.
Klahr, D., & Kotovsky, K. (Eds.). (1989). *Complex information processing: The impact of Herbert Simon*. Hillsdale, NJ: LEA.
Krauss, R. M., & Fussell, S. R. (1990). Mutual knowledge and communicative effectiveness. In J. Galegher, R. E. Kraut & C. Edigo (Eds.), *Intellectual teamwork: Social and technological foundations of cooperative work.* Hillsdale, NJ: Erlbaum.
Levine, J. M., & Thompson, L. (1996). Conflict in groups. In E. T. Higgins & A. Kruglanski (Eds.), *Social psychology: Handbook of basic principles.*
Maxwell, J. A. (2004). Causal explanation, qualitative research, and scientific inquiry in education. *Educational Researcher, March 3*, 3-11.
Morrow, R. A., & Brown, D. B. (1994). Deconstructing the conventional discourse of methodology *Critical Theory and Methodology* Thousand Oaks, CA: Sage Publication.
Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Pretice-Hall.
Paavola, S., Lipponen, L., & Hakkarainen, K. (2004). Models of innovative knowledge communities and three metaphors of learning. *Review of Educational Research, 74*(4), 557-576.
Pintrich, P. R. (1999). The role of motivation in prompting and sustaining self-regulated learning. *International Journal of Educational Research, 31*, 459-470.
Rosé, C. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning, 3*, 237-271.
Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. New York: Houghton Mifflin Company.
Stahl, G., Koschmann, T., & Suthers, D. D. (2006). Computer-supported collaborative learning: A historical perspective. In R. K. Sawyer (Ed.), *Cambridge handbook oh the learning sciences*. New York: Cambridge University Press.
Strijbos, J., Martens, R. L., Prins, F. J., & Jochems, W. M. G. (2006). Content analysis: What are they talking about? *Computers and Education, 46*, 29-48.
Suthers, D. D., Lund, K., Law, N., Rose, C., Teplovs, C., Chen, W., et al. (2011). *Towards Productive Multivocality in the Analysis of Collaborative Learning.* Paper to be presented at the CSCL 2011.
von Glaserfeld, E. (1987). Learning as a constructive activity. In C. Janvier (Ed.), *Problems of representation in the teaching and learning of mathematics* (pp. 3-38). Hillsdale, NJ: Erlbaum.
Weinberger, A., Stegmann, K., & Fischer, F. (2007). Knowledge convergence in collaborative learning: concepts and assessment. *Learning and Instruction, 17*, 416-426.
Yanchar, S. C., & Williams, D. D. (2006). Reconsidering the compatibility thesis and electicism: Five proposed guidelines for method use. *Educational Researcher, 35*(9), 3-12.

Acknowledgments

This research was in part funded by the National Research Foundation of Korea (NRF) (Grant No. 2009-0068919) awarded to the first author and a Rutgers Faculty Research Council grant to the second author.

Technology-mediated Reflection and Peer-exchange: Supports for Teacher Professional Development Communities

Cheryl Ann Madeira, James D. Slotta, University of Toronto, 252 Bloor Street West, Toronto, ON, Canada
Email: cheryl.madeira@utoronto.ca, jslotta@oise.utoronto.ca

Abstract: This design-based research study investigates the development of pedagogical content knowledge (PCK) among nine teacher-participants in three design phases. PCK is a particular type of teacher knowledge that addresses not only the teacher's understanding of the content to be instructed, but also ways of how to teach that content effectively. This paper offers a detailed perspective on how teachers developed PCK with their engagement in lesson planning and enactment of a project-based technology-enhanced lesson in relationship to their actions in the class. The study includes two specific interventions designed to enhance teachers' development of PCK: (1) scaffolded reflection, and (2) peer-exchange. The findings presented in this paper focus on the impact of the peer exchange intervention, which was achieved through an online community, combined with face-to-face meetings. Results demonstrate a positive impact of such exchanges on the quality of teachers' lesson designs and enactments.

Introduction and Objectives

How can teachers be supported as they develop their expertise and craft knowledge? Judith Warren Little (2002, p 2) comments, "Long-term observers of educational innovation and school reform have argued that reform might more productively be seen as a problem of learning than as a problem of implementation. That is, the progress of reform appears to reset in crucial ways on the capacity of teachers, both individually and collectively." Often, teacher learning happens in workshops where the topics are selected in a top-down fashion by school boards and ministries of education and lack any explicit connections to what teachers may already know, or any clear approach to engaging them in learning about their practice. In recent years, many educational and cognitive researchers have identified the pivotal role that teacher knowledge and expertise plays in student learning within an inquiry-based environment (Davis & Krajcik, 2005; Driver, Asoko, Leach, Mortimer, & Scott, 1994; Fishman, Marx, Best, & Tal, 2003). How can CSCL technologies and methods support their exchange of ideas, strategies and artifacts about their practice? This paper presents a longitudinal study of a technology-enhanced environment that supported teachers in the online exchange of their ideas and experiences, strengthening their learning of complex practices and their development of community. This design-based study investigated a teacher professional development community and the role that scaffolded reflections and peer-exchange has in teachers' growth of professional knowledge and practices.

Traditionally, teachers have tended to learn through isolated and informal circumstances, disconnected from dialogue or discussion with peers. However, in the Web 2.0 era, social exchanges play a more central role in learning, and technology has changed the ways in which learners make sense of the world (Collins & Halverson, 2010). The research literature identifies two important mechanisms for teacher knowledge development – namely reflection and peer exchange. Reflection is a cognitive process that helps learners to construct their own coherent understandings (Davis, 2006; Linn & Slotta, 2006). Through peer exchange, teachers can share ideas and experiences that lead to insights they may not otherwise have recognized (Davis, Smithey, & Petish, 2004). Technology can offer new ways for teachers to reflect on their practices as they engage in complex activities (Collins & Halverson, 2010) and new ways to interact within a community of peers (Slotta & Linn, 2009) as it provides a temporal space for the exchange of lesson plans, enactment strategies and epistemological understandings about student knowledge and learning.

This research examines a computer-supported teacher community where members co-designed, enacted, and revised project-based, technology-enhanced science lessons. The teachers were scaffolded in online and face-to-face reflections and exchanges that were designed to connect directly to their relevant professional practices (i.e., lesson planning and lesson enactment). The impact of scaffolded reflections was examined during three stages: (1) teacher lesson planning, (2) enactment, and (3) revision of lesson. The impact of peer exchange was investigated by establishing a co-design and discussion process for lesson planning, enactment and revision. Using a wiki layer and accompanying Web site, these two interventions (reflection and peer exchange) were studied through multiple design iterations. This paper provides a rich description of how individual teachers developed understandings within this context, contributing to our understanding of situated collaborative learning, and offering insight to models of teacher professional development.

Theoretical Foundations

Teacher Knowledge and Professional Development

Teachers' Pedagogical Content Knowledge (PCK), critical to the successful adoption of any innovative methods or materials, is defined as the blend of the their understanding of content within subject domain, the epistemological characteristics of learning within their subject domain, and the specific pedagogical practices and characteristics of their subject domain (Shulman, 1986, 1987; Borko & Putnam, 2002; Gess-Newsome, 1999; Loughran, Berry, & Mulhall, 2006). While the tacit nature of PCK, as connected to complex teacher actions within the classroom, hinders efforts to assess or evaluate it, there are several factors that are thought to influence its development, including teacher lesson planning (Magnusson et al., 1999), and teacher reflection (Park & Oliver, 2008). Recently, another form of teacher knowledge, namely technology pedagogical content knowledge (TPCK), has been advanced to describe knowledge that enables the navigation through the complex relationships of content, pedagogical and technology knowledge (Koehler, Mishra & Yahya, 2007). Teacher knowledge can be difficult to measure in the course of any study of classroom practice or professional development. There are some metrics that have been used to document teacher perceptions and beliefs about their own teaching and instructional practice (Loucks-Horsley & Matsumoto, 1999; Stuessy & Metty, 2007). Across a wide spectrum of research on teaching practices, two kinds of interventions have been seen are seen as effective in both monitoring and influencing teacher learning: reflection and peer-exchange (Slota & Linn, 2009). The next two sections of the paper provide a brief review of such research, which sets the context for our own study of these two important interventions.

Scaffolded Reflection

A substantial body of professional development literature has focused on reflective practice as a means of enabling the growth or development of teacher knowledge (Barnett & Coate, 2005). Schön (1983, 1987) offered a theoretical perspective about the growth of teacher knowledge during the course of professional practices, where the silent reflections occurring within the context of classroom activities make knowledge construction tacit. For example, if something catches a teacher's attention during class (e.g., a student misinterpretation of a concept), the teacher may recognize that moment as being significant and ponder on possible changes for the next class or the next iteration of the lesson. While such tacit reflections are likely an integral part of many teachers' practice (McNamara, 1990), limited time is provided for teachers to engage in deliberate reflections that are explicitly linked to classroom events, such as in reflecting about why some lessons are successful and why some are not (Raines & Shadow, 1995). This study recognizes the limitations in reflective practice and uses technology to scaffold and structure reflections that target various teaching practices of planning, enactment and revising of lessons. This type of reflective practice draws on an understanding of scaffolding for reflection, a general process where learners are supported with technology-based prompts or tools that enable them to achieve knowledge that might otherwise be ignored (Slotta & Linn, 2009; Fishman et al., 2003).

Teacher Communities and Peer-Exchange

The notion of a peer community suggests opportunities for social interaction, mutual dependence, and group identification, with (in the case of teachers) possible connections to classroom events (Westheimer & Kahne, 1993). Such a structure, should it ever be achieved, could be seen as reflecting Lave and Wenger's (1991) notion of a Community of Practice, in which participants' knowledge of practice is rehearsed both explicitly and implicitly. Palinscar et al. (1998) have suggested that if learning, thinking, and knowledge construction are inextricably connected to social contexts, then providing opportunities for teachers to engage in meaningful deliberation, interaction, and reflection within a community of peers would help to nurture their practice and professional development.

Following other approaches to establishing online communities, such as MERLOT (McMartin, 1997), Tapped In (Schlager et al., 2004) and MSPnet Hub (Falk & Drayton, 2004), this study defined online community as a persistent space for teachers to share approaches and interact with peers, serving not only as a digital resource, but a virtual space for collaborations (i.e., in lesson planning) and reflection (i.e., regarding lesson enactment). Participation or engagement within such a community can be measured in terms of contributions of materials (e.g., lesson plans, reflections or comments) and engagement in peer exchange (e.g., adding comments to one another's postings). Such online resources can help teachers to connect with peers across physical distances and asynchronously, allowing access for a wide range of participants. In the present study, the community consisted of a wiki, with an accompanying website, providing a space for persistent and continuous knowledge for the teachers to draw upon.

Methodology

This three-year, iterative, design-based study investigated teacher professional development as it occurs in the rich context of a curriculum-design community, where nine science teachers (N=9) each designed, enacted, and revised a technology-enhanced project-based lesson. The lessons were designed according to a generic set of characteristics for Project-Based Learning (Laffey et al., 1998; Blumenfeld et al., 1991) and used various technologies including productivity software (e.g., Microsoft Office), visualization tools (e.g., Inspiration) social technologies (e.g., wikis or blogs) and interactive learning environments (e.g., WISE: The Web-based Inquiry Science Environment). This study focused on the role of two primary forms of intervention in teacher's development of PCK: reflections and peer-exchange within a community. Iteration 1 included scaffolded reflection and limited community. In Iteration 2, design improvements included improved scaffolded reflection (more deliberate to project-based learning and content), and introduced peer-exchange.

Participants

There were nine science teacher participants (*N*=9) with a range of experience and disciplinary expertise (i.e., physics, biology, chemistry, or general science). Figure 1 illustrates participants' teaching experiences and subject expertise. They were selected based on their expressed interest and content knowledge, and were surveyed about their initial understandings of project-based instruction and technology. Teachers came from 5 different schools located in a large urban city in North America and had access to a fairly wide range of technology supports (i.e., as provided by their schools). The teachers and research-mentor formed a co-design team. The research-mentor was doctoral student who also had 17 years of experience teaching in secondary science classes.

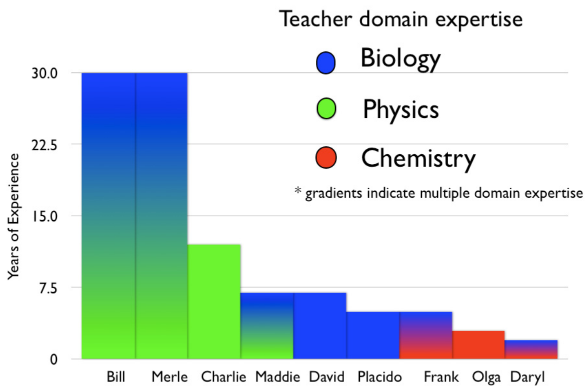


Figure 1. Teacher Participants - Years of Experience and Subject Expertise.

Materials

Pre-survey and Interview: In order to establish a measure of teachers' background and pedagogical content knowledge, a pre-survey was administered to all teachers, followed by an interview for purposes of clarification and to orient the teacher to the study. The following questions were among those given to teachers before starting the research project: (1) What are some of your best learning experiences and why do you think they were important?; (2) What are some of your visions within your science classroom?; (3) What are some of your previous project-based lessons that you have conducted?

Lesson Plan Template: A wiki site was designed with specific scaffolding categories for the teacher to design their new technology-enhanced project-based lesson. Some sample categories were (1) "Determining Topic", (2) "Challenges for students in science topic?", and (3) "How can technology help?".

Teacher Lesson Designs: Teachers co-designed a technology-enhanced project-based lesson using the scaffolded wiki-template with the research-mentor and their peers in the community. These lessons targeted a project-based learning pedagogy and include a variety of technology tools and materials. The lessons themselves were a focus for analysis. The table below provides an overview of the basic science topics and technology-enhanced project-based lesson themes selected by each teacher.

Table 1: Technology-enhanced project-based lessons.

Teacher Participant	Grade Level	Science Concept	Technology-enhanced PB Lessons
Bill	10	Chemistry Indicators	Photo Journal/Wiki/Poster of Experiment
Merle	12	Physics – interdisciplinary; several concepts	Wiki/Podcast Walking Tour of Science Exhibits
Charlie	11	Physics – Sound, Electromagnetism, Newton's Laws	Wiki/Video/Experiment Segment of concept
Maddie	11	Physics – Newton's Laws, Renewable Energy	Wiki/Video/Experiment Segment of concept
Frank	7/8	Ecology – Wolves management program	Wiki/WISE/Video Wolves Ecology Commercial
Placido	7/8	Earth and Space – Earth's Crust project	Wiki//Video/Powerpoint Presentation
Alex	7/8	Earth and Space – Earth's Crust project	Wiki//Video/ Powerpoint Presentation
Olga	12	Chemistry – interdisciplinary; several concepts	Wiki/Poster/Debate-Presentations
Daryl	9	Biology – Reproduction; Physics – Reusable Energy	Wiki/ Powerpoint Presentation

Reflection Questions: A wiki site and specific scaffolded reflection prompts were designed to capture teachers' understandings about pedagogical content. Reflection questions were asked at three different points in each teacher's lesson planning and enactment process:

- Prior to the lesson planning: (1) What are the goals of your lesson? (2) What are your thoughts about the student ideas? (3) What are some of the key elements in your project-based design?
- After the lesson planning: (1) Where do you think the students will be challenged? (2) What will you try to do with your time, during the lesson? (3) What are some of the follow up concepts that require set up during lesson?
- After the lesson enactment: (1) Did you find the students had more or less difficulty than you expected? (2) What is one change or addition you would like to put into place for next time? (3) What was one advantage in using the technology within the project-based activity?

Community: The peer community within this project includes both offline and online components. The teachers had periodic community meetings where they exchanged ideas and shared their stories about the project-based enactment, which served to establish a personal relationship between community members. The online component of the teacher community consisted of a website and a wiki site, developed to collect personal statements from teachers about their background and philosophy, as well as to collect details of lesson plans, and all reflections.

The online community supported peer exchanges, reviewed lesson plans and fostered discussions about the enactments. Upon completing the lesson plan an update of the 'lessons learned' and the 'things I hope to add to the lesson next time' was added to the wiki lesson page. Teachers in a community were asked to connect to their peers by asking questions and commenting to this additional wiki page. The wiki site for lesson design and reflection played a significant role in supporting socially constructed knowledge. It enabled teachers and the mentor to make their knowledge visible for themselves and all members of the community. These on-line artefacts became an assets for reference by all members of the community. Figure 2 displays the Website interface of the on-line community (left) which linked to Figure 3, the wiki space (right).

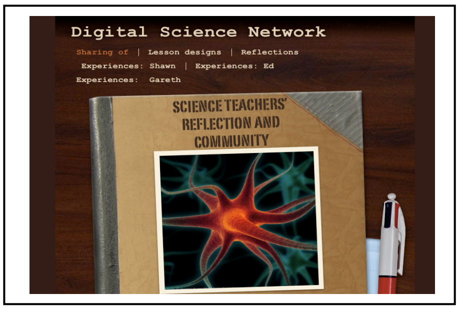


Figure 2. Teacher Web Interface.

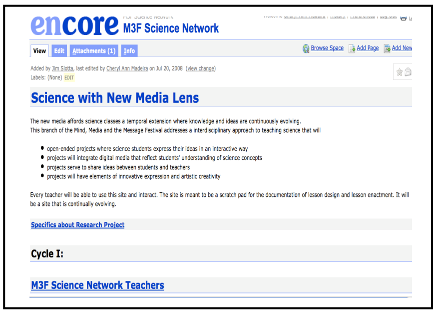


Figure 3. Teacher Wiki-community Site.

Procedure

This paper reports on nine science teachers' development of PCK and their experiences during 3 design phases of curriculum design and enactment. There were three main teacher activities that occurred within each iteration: (1) Lesson design, (2) Classroom enactment, and (3) Revision of lesson design. Data sources include teacher surveys, interview questions, wiki lesson plans, scaffolded reflections (captured in a wiki), videotaped classroom enactments, field notes, student artifacts and wiki responses, teacher peer-exchanges (on a wiki and in group meetings).

The first design phase of the study included four teachers (n = 4) who worked individually with the researcher-mentor to co-design, enact and reflect on a science lesson. The second design phase added five more science teachers (n=5), increasing the community to a total of nine teachers and one researcher. Consistent with the design-based research paradigm, improvements were made on the reflective prompts (i.e., connecting reflection prompts to lesson planning and enactment) and community elements (face-to-face and online), which marked the beginning of the second design phase. The third phase of the design study continued to refine the scaffolded reflection and community exchange connecting teacher activities of lesson planning and enactment deeply with student prior knowledge, project-based learning and technology implementation.

Analysis and Findings

Community Exchange of Content

The online teacher community helped to make teacher ideas and actions more visible, accessible, and persistent, fostering scaffolded discussions about how to revise lesson plans and helping to focus teachers' thinking on issues of student learning. Figure 5 shows a screen capture of one teacher's (Charlie) lesson plan and Figure 6 shows a sample of a student artifact that he posted for his peers: a digital resource he had created, consisting of a wiki page to support students' planning for a video project. Such sharing allowed other teachers to see how their peer had set up lesson plans and how his lesson had taken form in students' hands. The community was able to read Charlie's lesson plan and his reflections about it, and add their own comments (Figure 7). The teacher-participants could follow the changes that Charlie made to this lesson from Iteration 1 to Iteration 2, as well as his rationale for those changes, and see the resulting impact on student work.

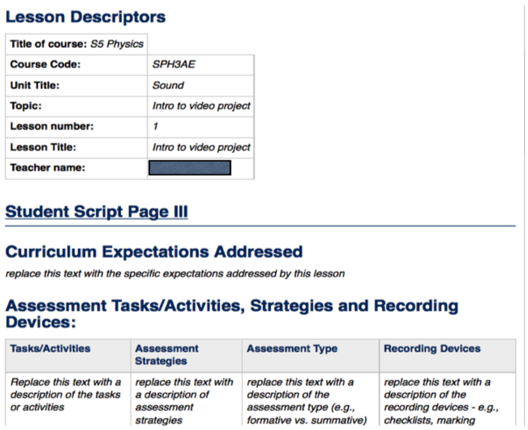


Figure 5. Charlie's Lesson Plan—Iteration 1.

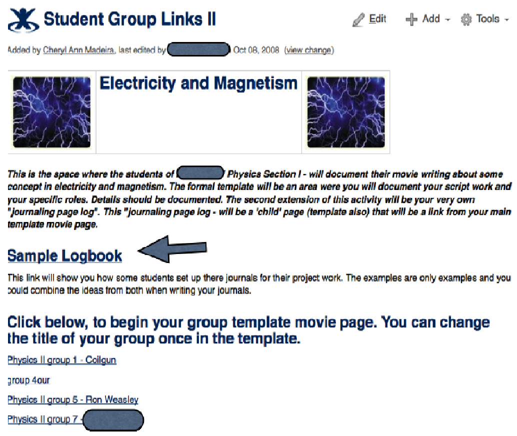


Figure 6. Student Wiki-template Iteration 2; with Lesson Design Improvements Including Student Reflection.

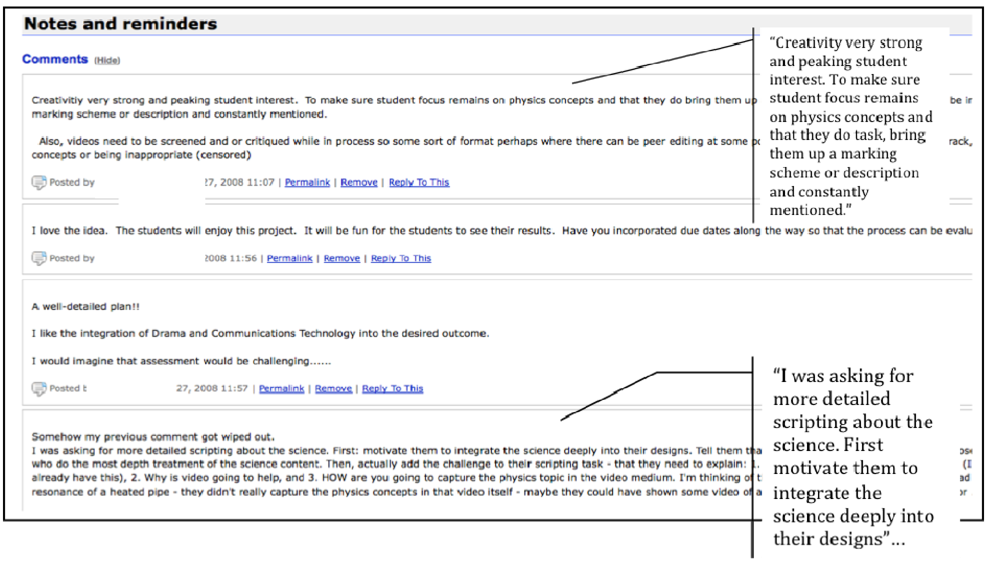


Figure 7. Community Reflections on Lesson Plans (comm27/08/08).

In the community meetings, Charlie described how his video project had allowed students to express their own understanding of physics topics, but also detailed where the lesson design activity had fallen short in engaging them. As he planned his second iteration, he published his improvements to the community Website. He introduced wiki logbooks to help students make sure that they documented what they were doing throughout the process. After enacting this version twice, Charlie made substantive revisions in Iteration 3, introducing improvements to his wiki support materials (for students) to help make the lesson design more visible to students as well as to his peers in the community.

Teacher Engagement

Teachers who had entered this project with limited understandings of technology were able to experience, through their own engagement within the online environment, how the use of the various technology tools could have potential benefits within their classrooms. The use of the Website as a portal for the community allowed teachers to easily follow lesson plans, reflections, and video clips of their peers' enactment. It offered an effective means of organizing teachers' artifacts and reflections and supporting their participation.

Teachers' participation within both the online and face-to-face community elements was coded from 1 (lowest) to 4 (highest) based on a rubric that examined three aspects of community engagement: (1) Did the participants provide or find strategies for project-based learning? (2) Did the participants provide or find effective uses of technology? (3) Did the participants provide or find strategies for promoting student inquiry and learning? Other areas such as content sharing and reflection were seen as implicitly fitting within those three elements. Thus, teachers scored highly on this rubric when they had provided lesson plans and student artifacts, discussed their strengths and limitations with peers, reflected on the challenges and strengths of their enactments, and shared their assessment and lesson revisions. Teachers scored lower when they contributed little to the reflections or comments within the online space or face-to-face meetings. A score of 1 was awarded as the lowest possible measure, acknowledging that all participants did have some level of social engagement - at least with the mentor. The face-to-face community involvement was scored and ranked in a manner similar to the scoring used for online participation, providing a simple qualitative measure of participation in meetings or online tasks. Table 2 provides all scores and gives a brief explanation of the teachers' participation in online and face-to-face activities.

Table 2: Community engagement (averaged time of involvement with the study).

Teacher-Participant	Face-to-face Engagement		Online Engagement		Total Score	Comment about Participant Engagement in Community (evaluated the number of times attending and items contributed in the community)
	High	Low	High	Low		
Charlie	4		4		8	Attended every community meeting, and always participated online; shared artefacts, assessments, and reflections
Merle	4		4		8	Attended every community meeting, and always participated online; shared artefacts, assessments, and reflections
Maddie	4		4		8	Attended every community meeting, and always participated online; shared artefacts, assessments, and reflections
Olga	3		3		6	Attended 3 community meetings, and always participated online; shared artefacts, assessments, and reflections
Alex	3		3		6	Attended 3 community meetings, participated online, shared artefacts and reflections and participation online; shared some assessments
Frank		2		2	4	Attended only 1st community meeting, and limited participation online; shared artefacts, assessments, and reflections (no job next term)
Bill		1		1	2	Attended only 1 community meeting, video taped himself for one community meeting, limited participation online; limited sharing
Placido		1		1	2	Attended only 1 community meeting, very limited participation online; shared artefacts, assessments, and reflections
Daryl		1		1	2	Never attended community meetings, no participation online; shared artefacts, assessments, and reflections (before online community).

Figure 8 positions places teachers along a two-dimensional grid of online and face-to-face participation in the study. Each teacher's position is obtained from Table 2, and represents the level of participation observed in each dimension. Further, an overall engagement score for participation or engagement in community can be obtained by summing the two dimensions. It is interesting to note that those teachers who scored low in one form of community (e.g., online) also scored low on the other (e.g., face-to-face). In other words, teachers either appear in the lower left or the upper right quadrants.

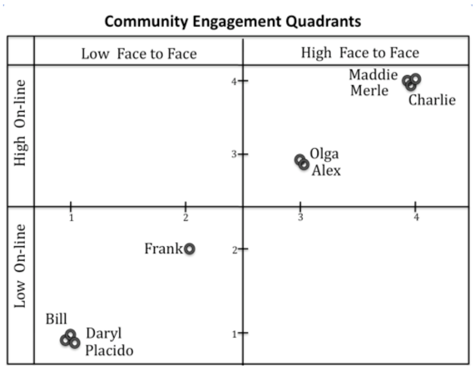


Figure 8. Online and Face-to-Face Engagement.

While the two-dimensional grid may seem arbitrary, it offers a depiction of how well teachers engaged in the community intervention. This measure can also be used to validate the scoring of the Community elements in the next section. The data sources for the Community elements were from scaffolded wiki reflections, interviews, and field notes. These could be viewed as more subjective and had a quality score. The community engagement score was a numerical value of participation. For the purposes of this paper, only Community element with lesson planning will be described.

Community Impact on Lesson Planning

To evaluate the impact of the community intervention, we developed a coding scheme and scoring rubric that captured teachers' participation relative to the available community supports (i.e., in any given design phase of the study). Peer exchange influenced lesson planning, as teachers became aware of other teacher-participants' projects and discussed lesson planning during community meetings. They were able to access lesson plans from the wiki site, observe student artefacts, and read their peers' reflections about enacting the lessons. The Community elements within the Lesson Planning rubric focused on explicit interactions with mentor and peers. Indeed, one element in the rubric was called "Interaction with mentor," and the other was called "Influence of community" (see Figure 9). Each of these elements was scored according to the rubric on a 1 (lowest) to 3 (highest) scale. The following ranked excerpts illustrate the different coding scores for the element "Interaction with mentor."

Rubric: Score of 1. Response indicates limited mentor influences. Answer has limited use of mentor for technology.

Excerpt: *[Participant wrote not applicable] N/A (Daryl 301107)*

Rubric: Score of 2. Response indicates some use of mentor influences. Answer has some use of mentor for technology.

Excerpt: Sharing ideas[with technology], cooperation and working together, team building, working together[with wiki], and trusting that everyone (teachers and mentors) participates and puts an effort. (Frank 280508)

Rubric: Score of 3. Response indicates strong use of mentor influences. Answer has strong use of mentor for technology.

Excerpt: *I would not have been able to consider the changes to the lesson or even trying this video project-based lesson without the community, without you* [researcher-mentor], *so it has changed me and my perspectives giving me things to think about…student-peer exchange and scaffolding their understanding. (Charlie 291008)*

The average Community score was tallied for all teachers across their two iterations. Figures 9 and 10 show that this community score correlates with the lesson planning overall score for Iteration 1 and Iteration 2, respectively. These correlations were both marginally significant, with $F < 0.05$, and $F < 0.07$, respectively. This provides some evidence of a link between teachers' community influence during lesson planning and the quality of their lesson design process. This suggests that teachers can improve in lesson planning by engaging in peer exchange. Given that reflection was also seen to impact on lesson planning, the combined effects of reflection and community should be an even more effective means of professional development.

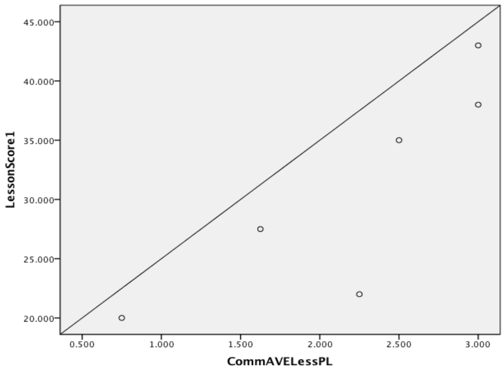


Figure 9. Correlation of Lesson Planning Score and Community Average Score Iteration I.

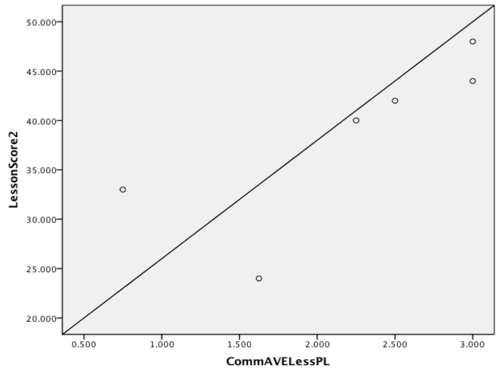


Figure 10. Correlation of Lesson Planning Score and Community Average Score Iteration 2.

Conclusions

Through the use of CSCL tools like scaffolded reflections and peer exchanges, teachers in this study gained a deeper understanding about project-based instruction and translated it into practice within their classrooms. The technologies used within this study offered ways for teachers to exchange their reflections, lesson plans, strategies and student artefacts. The technology also supported the design-based methodology of this study, as teachers were able to progress between iterations in reference to their wiki-based lesson plans and templates. This allowed for more informed improvements to the lesson plans, and also allowed for the iterative refinement of the scaffolded reflections and peer-exchange prompts and activities. This research provides support for a model for professional development that engages teachers through active reflection and exchange of ideas and experiences with peers, supporting their understanding of new practices and their development of peer communities.

References

Barnett, R., & Coate, K. (2004). *Engaging the curriculum in higher education*. Maidenhead, England: McGraw-Hill/Open University Press.

Blumenfeld, P., Soloway, E., Marx, R., Krajcik, J. S., Guzdial, M., & Palincsar, A. (1991). Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational Psychologist, 26*(3/4)*,* 369–398.

Borko, H., & Putnam, R. T. (1996). Learning to teach. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 673–708). New York, NY: Macmillan.

Collins, A. & Halverson, R. (2010). *Rethinking Education in the Age of Technology: The Digital Revolution and Schooling in America.* New York: Teacher College Press.

Davis, E. A., & Krajcik, J. S. (2005). Designing educative curriculum materials to promote teacher learning. *Educational Researcher, 34*(3)*,* 3–14.

Driver, R., Asoko, H., Leach, J., Mortimer, E., & Scott, P. (1994). Constructing scientific knowledge in the classroom. *Educational Researcher, 23*(7)*,* 5–12.

Gess-Newsome, J. (1999). Pedagogical content knowledge: An introduction and orientation. In J. Gess-Newsome & N. G. Lederman (Eds.), *PCK and science education* (pp. 3–17). Dordrecht, the Netherlands: Kluwer Academic.

Falk, J., & Drayton, B. (2004). State testing and inquiry based science: Are they complementary or competing reforms? *Journal of Educational Change, 54*(4)*,* 345–387.

Fishman, B., Marx, R., Best, S., & Tal, R. (2003). Linking teacher and student learning to improve professional development in systemic reform. *Teaching and Teacher Education, 19*(6)*,* 643–658.

Koehler, M. J., Mishra, P., & Yahya, K. (2007). Tracing the development of teacher knowledge in a design seminar: Integrating content pedagogy & technology. *Computers & Education, 49,* 740–762.

Laffey, J., Tupper, T., Musser, D., & Wedman, J. (1998). A computer-mediated support system for project-based learning. *Educational Technology Research and Development, 46*(1)*,* 73–86.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York, NY: Cambridge University Press.

Little, J.W. (2002) Locating learning in teachers' professional community: Opening up problems of analysis in records of everyday work. *Teaching and Teacher Education 18* (8), 917-946.

Loucks-Horsley, S., & Matsumoto, C. (1999). Research on professional development for teachers of mathematics and science: The state of the scene. *School Science and Mathematics, 99*(5)*,* 258–271.

Loughran, J., Berry, A., & Mulhall, P. (2006). *Understanding and developing science teachers' pedagogical content knowledge*. Rotterdam, Netherlands: Sense.

Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature, sources, and development of pedagogical content knowledge for science teaching. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge: The construct and its implications for science education* (pp. 95–132). Dordrecht, the Netherlands: Kluwer Academic.

McMartin, F. (2008). Open education through open educational resources. In T. Iiyoshi & M. S. V. Kumar (Eds.), *Opening up education: The collective advancement of education through open technology, open content, and open knowledge* (pp. 135–178). Cambridge, MA: MIT Press.

McNamara, D. (1990). Research on teachers' thinking: Its contribution to educating student teachers to think critically. *Journal of Education for Teaching, 16*(2), 147-160.

Palincsar, A. S., Magnusson, S., Marano, N., Ford, D., & Brown, N. (1998). Designing a community of practice: Principles and practices of the GIsML community. *Teaching and Teacher Education, 14*(1)*,* 5–19.

Park, S., & Oliver, J. S. (2008). Revisiting the conceptualization of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. *Research in Science Education, 38,* 261–284.

Raines, P., & Shadiow, L. (1995). Reflection and teaching: The challenge of thinking beyond the doing. *The Clearing House, 68*(5)*,* 271–274.

Schlager, M. S., & Fusco, J. (2004). Teacher professional development, technology, and communities of practice: Are we putting the cart before the horse? In S. Barab, R. Kling, & J. Gray (Eds.), *Designing virtual communities in the service of learning* (pp. 120–153). New York, NY: Cambridge University Press.

Schön, D. (1983). *The reflective practitioner*. New York, NY: Basic Books.

Schön, D. (1987). *Educating the reflective practitioner*. San Francisco, CA: Jossey-Bass.

Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(1)*,* 4–14.

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57(1), 1–22, 57*(1)*,* 1–22.

Stuessy, C. L., & Metty, J. S. (2007). The learning research cycle: Bridging research and practice. *Journal of Science Teacher Education, 18,* 725–750.

Slotta, J.D. & Linn, M.C. (2009). *WISE science: Web-based inquiry in the classroom*. New York, NY: Teachers College Press.

Westheimer, J., & Kahne, J. (1993). Building school communities: An experience-based model. *Phi Delta Kappan, 75*(4)*,* 324–328.

AUTHOR INDEX