Proposals for Research

Gerry Stahl
Gerry Stahl's Assembled Texts

- Marx and Heidegger
- Tacit and Explicit Understanding in Computer Support
- Group Cognition: Computer Support for Building Collaborative Knowledge
- Studying Virtual Math Teams
- Translating Euclid: Designing a Human-Centered Mathematics.
- Constructing Dynamic Triangles Together: The Development of Mathematical Group Cognition
- Essays in Social Philosophy
- Essays in Personalizable Software
- Essays in Computer-Supported Collaborative Learning
- Essays in Group-Cognitive Science
- Essays in Philosophy of Group Cognition
- Essays in Online Mathematics Interaction
- Essays in Collaborative Dynamic Geometry
- Adventures in Dynamic Geometry
- Global Introduction to CSCL
- Editorial Introductions to ijCSCL
- Proposals for Research
- Overview and Autobiographical Essays
- Theoretical Investigations
- Works of 3-D Form
Gerry Stahl’s assembled texts volume #17

Proposals for Research

Gerry Stahl

2015
Introduction

The purpose of this volume is to share the proposals that I have made for research, including first of all those that have been funded and have allowed me to engage in an active research agenda, both at the University of Colorado in Boulder as a Research Professor and at Drexel University in Philadelphia as an Associate Professor and Full Research Professor.

I have also included a number of proposals that I felt should have been funded; these document ideas that I was working on at the time they were written, but ultimately represent roads not taken. They were modest (more or less) proposals for promising, but unfulfilled, research potentials. Perhaps they document stages in the development of my thinking not otherwise visible; perhaps they will inspire a reader to pursue an otherwise forgotten trail of inquiry.

Writing effective, competitive grant proposals is a delicate business. First, one has to conceive of a program of research that one would like to undertake and that is reasonable to attempt under the proposed conditions. Then, one must convince the funding source and their reviewers that the funding proposal should be accepted. This must be accomplished with a written document of restricted form, content and length.

Preparing a proposal is a challenging writing task, requiring project planning, persuasive presentation and organized narrative. In many ways it is like writing a professional research report for a journal, such as one would compose near the end of the prospective funded project, but it needs to include more than just the concept, theory, literature review and analysis. It also needs to demonstrate why the person or group proposing is the right one to do the job and detail how the work is expected to be accomplished with the requested resources. In this publication, I only include the Proposal Summary, Proposal Description and Proposal References. The details of personnel and budget are too specific to be of interest to the reader.

I attribute my success in grantsmanship to a number of stages in my life. Most likely, I honed my natural argumentation tendencies through a decade of study of philosophy (Stahl, 2010a; 2010b). But this left my writing style
too abstruse for the practical world of grant funding. Once I had completed
my doctoral study of philosophy, I returned to the streets of Philadelphia as
a community organizer in the 1970s of the Great Society era of federal
funding. My first proposal was awarded a million dollar grant to a network
of neighborhood organizations to train unemployed residents in poor
neighborhoods to start energy conservation, recycling and home repair
projects. I later joined the Southwest Germantown Community
Development Corporation as community planner and brought in dozens of
federal, state, city and foundation grants over several years to support a local
credit union, an energy conservation organization and neighborhood projects
in youth employment, housing rehab and economic development. This
taught me not only proposal writing, but project management, especially
non-profit fund accounting and budgeting. Next, I provided technical
assistance to non-profit organizations throughout Philadelphia and started a
computerization service for them when the first personal computers came
along, developing custom accounting and service tracking software.

In 1989, I moved out West and studied computer science, artificial
intelligence and cognitive science in Boulder (Stahl, 2010c). I helped
writing proposals for the lab I was in and drafted the proposal that paid for
my post-doc position.

After graduation, I worked with a small research start-up, drafting SBIR
(federal small-business innovative research program) proposals for research
in collaboration with firms and government agencies. I worked as VP for
R&D, doing the programming for grants that were funded. The projects
were in collaboration with the Boulder Department of Education and with
the astronaut psychology group in NASA. Some of this research is reported
in Group Cognition.

I eventually became a Research Professor at the Institute of Cognitive
Science and the Department of Computer Science. This meant that I had to
raise my entire salary from grants, so I began writing proposals intensively.
While I was awarded some relatively small grants, I never succeeded in the
almost impossible job of supporting myself as a research professor.

I went to work at a CSCW lab in Germany for a year and then joined the
faculty of the College of Information Science and Technology (the iSchool)
at Drexel University. There, I met the people at the Math Forum at Drexel
and developed collaborations that resulted in successful grant proposals and
productive research. My grants raised over six million dollars to support the
VMT Project from 2003-2016.
The following pages are organized in retrospective chronology, divided in four Parts:

Part I. Grants awarded at Drexel University (2003-2010)

- “Computer-Supported Math Discourse Among Teachers and Students.” Supplementary award DRL-1448116 from the National Science Foundation Discovery Research K-12 (DR K-12) Program for $152,743 over 2 years on September 1, 2014. PI: Gerry Stahl; co-PI: Stephen Weimar. For programmer salary to develop VMT-mobile technology.

- “Computer-Supported Math Discourse Among Teachers and Students.” Supplementary award DRL-135021 from the National Science Foundation Discovery Research K-12 (DR K-12) Program for $120,000 over 3 years on September 1, 2013. PI: Gerry Stahl; co-PI: Stephen Weimar. For participant support of teacher stipends and student prizes.


• "Engaged Learning in Online Communities." Award SBE-0518477. Funded by the National Science Foundation Science of Learning Center Catalyst Program for $180,762 over 3 years on October 1, 2005. PI: Gerry Stahl; co-PIs: Sharon J Derry (Wisconsin); K. Ann Renninger (Swarthmore); Mary R Marlino (UCAR); Daniel D Suthers (Hawaii). Project description: GerryStahl.net/publications/proposals/slc2005.

• "IERI: Catalyzing & Nurturing Online Workgroups to Power Virtual Learning Communities." Award IERI 0325447. Funded by the National Science Foundation IERI Program for $2,300,00 over 5 years on September 1, 2003. PI: Gerry Stahl; co-PIs: Stephen Weimar and Wesley Shumar. Project description: GerryStahl.net/publications/proposals/itr2003

• "Collaboration Services for the Math Forum Digital Library." Award DUE 0333493. Funded by the National Science Foundation NSDL Services Program for $450,000 over 3 years on August 15, 2003. PI: Gerry Stahl; co-PIs: Stephen Weimar and Wesley Shumar. Project description and proposal reviews: GerryStahl.net/publications/proposals/nsdl2003

Part II. Other proposals at Drexel University (2003-2010)


• “Multidisciplinary Curriculum Improvement and Innovation Using Software Defined Radio.” Proposal to the National Science Foundation Course, Curriculum, and Laboratory Improvement (CCLI) Program (Phase I — Exploratory). Submitted for $200,000 over 2 years on May 21, 2009. PI: Kapil Dandekar (Drexel ECE); co-PI: Gerry Stahl (Drexel).


• "CDI-Type 1: Building a world of math discourse using a mix of platforms." Preliminary proposal to the National Science Foundation Cyber-Enabled Discovery and Innovation (CDI) Program. Submitted for $797,303 over 3 years on Jan. 8, 2008. PI: Werner Krandick (Department of Computer Science, Drexel University); co-PI: Gerry Stahl (IST, Drexel).


• "Optimizing Feedback for Eliciting Pedagogically Valuable Explanation in Collaborative Problem Solving." Proposal to the National Science
Proposals for Research


• "Engaged Learning in Online Communities." Proposal to the National Science Foundation Science of Learning Center Catalyst Program. Submitted for $180,762 over 1.5 years on January 14, 2005. Proposal 0518477: 
  GerryStahl.net/publications/proposals/engaged/description.pdf.

• "Interaction Math: An Informal Online Learning Collaboratory Led by the Math Forum @ Drexel." Proposal to the National Science Foundation Informal Science Education Program. Submitted for $2,933,126 over 5 years on January 6, 2005. PI: Gene Klotz (Math Forum); co-PIs: Gerry Stahl and Stephen Weimar. Proposal 0515544: 
  GerryStahl.net/publications/proposals/informal/description.pdf.

• "Studying Online Collaborative Learning at the Math Forum." Proposal 337162 to the National Science Foundation ROLE Program. PI: Gerry Stahl; co-PIs: Scott Robertson and Wesley Shumar. Submitted for $1,790,931 over 3 years on June 1, 2003. Proposal: 
  GerryStahl.net/publications/proposals/role2003

• "Collaboration Services for the Math Forum Digital Library." Proposal 333493 to the National Science Foundation NSDL Services Program. PI: Gerry Stahl; co-PIs: Stephen Weimar and Wesley Shumar. Submitted for $494,953 over 2 years on April 21, 2003. Proposal: 
  GerryStahl.net/publications/proposals/nsdl2003

• "Group Knowledge Construction in Digital Library Communities." Proposal to the National Science Foundation NSDL Targeted Research Program. Submitted for $498,748 over 2 years on April 21, 2003. PI: Scott Robertson; co-PIs: Gerry Stahl and Susan Weidenbeck. Proposal 0333471: 
  GerryStahl.net/publications/proposals/nsdl2003b

• "ITR: Catalyzing & Nurturing Online Workgroups to Power Virtual Learning Communities." Proposal to the National Science Foundation ITR Program. PI: Gerry Stahl; co-PIs: Stephen Weimar and Wesley Shumar. Submitted for $3,374,472 over 5 years on February 12, 2003. Proposal 0325447: 
  GerryStahl.net/publications/proposals/itr2003

• "Educational Online Communities for At-Risk Youth." Proposal to foundations. Written for $88,000 over 1 year in December 2002. Proposal: 
  GerryStahl.net/publications/proposals/nursing2003/nursing.doc
Part III. Grants awarded at the University of Colorado (1997-2001)


Part IV. Other proposals at the University of Colorado (1997-2001)

- “Enhancing collaborative learning among researchers, practitioners, and students at CSCL 2002” (co-PI with Gerhard Fischer & Hal Eden) $49,860; Sponsor: NSF. Proposal 124010:
- “New Media to Support Collaborative Knowledge-Building: Beyond Consumption and Chat” (Principal Investigator) Proposal to the Lab for New Media Strategy and Design. Submitted for $19,752 over 4 months


- "Collaborative Web-Based Tools for Learning to Integrate Scientific Results into Social Policy" (co-PI with Ray Habermann at NOAA) $89,338; Sponsor: NSF.


- "ROLE proposal: The Role of Computational Cognitive Artifacts in Collaborative Learning and Education" (Principal Investigator) Proposal 106950 to the National Science Foundation ROLE Program. Submitted for $970,971 over 3 years on December 1, 2000. Proposal: http://GerryStahl.net/publications/proposals/role/role.pdf.

- "ROLE Pre-proposal: The Role of Computational Cognitive Artifacts in Collaborative Learning and Education" (Principal Investigator) Proposal 96877 to the National Science Foundation ROLE Program. Submitted for $750,000 over 3 years on September 1, 2000. Encouraged


• "Collaborative Research on Knowledge-Building Environments: Growing a National and International Research Community for Distance Learning Information Technology" (Principal Investigator) Proposal 77095 to the National Science Foundation. Pre-proposal submitted for $2,700,000 over 5 years on January 5, 2000. Proposal: http://GerryStahl.net/publications/proposals/collab/collab.pdf.


Note

This book does not include pre-proposals or versions of proposals that were resubmitted. It also does not include collaborative proposals that were primarily written by colleagues. Digital versions of most of my academic funding proposals are available at: http://gerrystahl.net/research.

References


## Contents

Introduction .................................................................................................................. 5  
Contents ......................................................................................................................... 17

### Part I: Grants Awarded at Drexel University

<table>
<thead>
<tr>
<th>Grant</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>DR K-12</td>
<td>Computer-Supported Math Discourse Among Teachers and Students</td>
<td>23</td>
</tr>
<tr>
<td>ONR</td>
<td>Theories and Models of Group Cognition</td>
<td>60</td>
</tr>
<tr>
<td>ALT</td>
<td>Dynamic Support for Virtual Math Teams</td>
<td>94</td>
</tr>
<tr>
<td>SLC</td>
<td>Engaged Learning in Online Communities</td>
<td>128</td>
</tr>
<tr>
<td>ITR</td>
<td>Catalyzing &amp; Nurturing Online Workgroups to Power Virtual Learning Communities</td>
<td>160</td>
</tr>
<tr>
<td>NSDL</td>
<td>Collaboration Services for the Math Forum Digital Library</td>
<td>194</td>
</tr>
</tbody>
</table>

### Part II: Other Proposals at Drexel University

<table>
<thead>
<tr>
<th>Grant</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>DR K-12</td>
<td>Computer-Supported Math Discourse Among Teachers and Students</td>
<td>233</td>
</tr>
<tr>
<td>ROLE</td>
<td>Studying Online Collaborative Learning at the Math Forum</td>
<td>269</td>
</tr>
<tr>
<td>Foundations: Educational Online Communities for At-Risk Youth</td>
<td>309</td>
<td></td>
</tr>
</tbody>
</table>
Part III: Grants Awarded at the University of Colorado

New Media to Support Collaborative Knowledge Building: Beyond Consumption and Chat

Interoperability among Knowledge-Building Environments

Conceptual Frameworks and Computational Support for Organizational Memories and Organizational Learning

Allowing Learners to be Articulate: Incorporating Automated Text Evaluation into Collaborative Software Environments

Part IV: Other Proposals at the University of Colorado

Perspectives on Collaboration: A Micro-ethnographic Study of Computational Perspectives in Computer Support for Collaborative Knowledge Building at a Virtual Biology Laboratory

Information Technology for Distributed Collaborative Learning in a Virtual Biology Lab

The Role of Computational Cognitive Artifacts in Collaborative Learning and Education

Perspectives on Collaborative Knowledge Building

Collaborative Research on Knowledge-Building Environments: Growing a National and International Research Community for Distance Learning Information Technology

IT Support for Knowledge Building in Workgroups
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Models for Organizing Collaboration: Ways of Supporting Distributed Learning</td>
<td>592</td>
</tr>
<tr>
<td>POW! (Perspectives On the Web)</td>
<td>598</td>
</tr>
<tr>
<td>Perspectives On the Web (POW!)</td>
<td>608</td>
</tr>
<tr>
<td>The Research CyberStudio: Supporting Researchers as LifeLong Learners</td>
<td>617</td>
</tr>
</tbody>
</table>
Part I: Grants Awarded at Drexel University
DR K-12: Computer-Supported Math Discourse Among Teachers and Students

This full research-and-development project designs, develops and tests an interrelated system of technological, pedagogical and analytic components to provide a range of opportunities for middle- and high-school students to engage in significant mathematical discourse (DR K-12 challenge 2); it catalyzes and supports these opportunities by enhancing the ability of in-service teachers to engage in, appreciate and foster math-problem-exploration and math-discourse skills in their students (DR K-12 challenge 3). The project addresses the core STEM discipline of mathematics by motivating the identification, comprehension and enjoyment of mathematical discourse skills through socially interactive, collaborative learning experiences involving pedagogically organized series of stimulating, skill-appropriate problems using computer-based visualization/exploration and small-group math-problem discussion.

The project’s design-based-research approach crafts a socio-technical educational model to provide a comprehensive, practical package of tools and techniques for classroom teachers and students, which integrates and refines a number of mutually supportive components: (a) Innovative technology: A custom, open-source virtual learning environment that integrates synchronous and asynchronous media with the first multi-user dynamic-math-visualization application. (b) Curricular resources: Problem-based learning topics in specific areas of mathematics designed to help teachers tune rich math problems to local texts or curriculum and to guide student exploration. (c) In-service teacher professional development: Practicing teachers in online masters programs are mentored to understand and model the innovative technologies and pedagogies by doing collaborative problem posing/exploring/solving and engaging in collaborative reflection on the math discourse in their logged interactions. (d) Middle- and high-school students: The teachers introduce the model, technology and resources into their classrooms.

The project builds on and integrates previous work of the PIs, including: the discourse-analysis-based theory of group cognition (Stahl,
2006); the Virtual Math Teams learning environment developed, analyzed and evaluated in (Stahl, 2009b); curricular materials and dynamic math visualization software of GeoGebra, adapted to flexible multi-user collaborative learning; online professional development and online mentoring of in-service math teachers at the Math Forum and at the Drexel and Rutgers-Newark schools of education; and the adaptation of conversation analysis to text-based chat interaction analysis, designed to highlight how collaborative problem solving or group knowledge building takes place. The project adapts components that have been explored, prototyped, or piloted by the PIs to classroom use. Project key personnel and Advisory Committee members bring expertise and experience in educational software R&D; math problem-set adaptation, dissemination and mentoring; in-service math teacher training; online math resources, collaborative learning, problem-based learning and dynamic math; design-based educational research management and evaluation; theory of knowledge building in small groups and in online communities. They also bring opportunities for national deployment and scaling up.

**Intellectual merit.** This project integrates leading-edge cyber-learning-environment technology incorporating innovative collaborative math exploration tools with educational approaches based on current directions in the learning sciences. It approaches this through a systematic iterative process of co-evolving the technology and curricular resources in the context of engaging, reflective collaborative-learning experiences of significant mathematical discourse by in-service teachers and their students. It thereby advances theory, technology and practice within real-world educational settings to forge a coherent research-based approach to math education appropriate to today’s challenges and potentials.

**Broader impact.** The project designs, tests, integrates, evaluates and disseminates technology, curricular resources, pedagogical methods and analytic tools for use in math-teacher professional-development programs, classrooms of math students, home-schooling networks, online schools and the Math Forum community (over three million visits per month). Project results will support the use of math exploration technology within collaborative math-discourse approaches at diverse schools nationally through their spread to in-service teacher-training programs and services—bringing practical cyber-learning of math to at-risk and isolated math students. It documents the potential impact on both teachers and students of this computer-supported math-discourse approach quantitatively and qualitatively.
Project Description

Mathematics education in the future faces enormous opportunities from the availability of ubiquitous digital networks, from innovative educational approaches based on theories of collaborative learning and from rich resources for interactive, online, dynamic math exploration. The fact that more and more teachers and students are learning online—with distance education, online masters programs, home schooling, online high schools, etc.—makes the incorporation of virtual collaborative learning environments a natural trend. A major issue in realizing these opportunities on a broad scale in schools is empowering teachers to appreciate and engage in the new approaches, and supporting them with appropriate tools, models and resources for practical instructional usage.

This project therefore proposes to develop a model of professional development and a suite of supports for math teachers. It will design, test, evaluate and refine a virtual learning environment that integrates synchronous and asynchronous media with an innovative multi-user version of a dynamic math visualization and exploration toolbox. Online teams of in-service teachers will be introduced to the collaborative exploration of Common Core State Standards-based math topics in this environment. They will then be guided in reflection on their own team’s discourse with the use of chat-replaying tools. As they become familiar with the use of the technology and with the nature of collaborative math discourse, some of the trained teachers will mentor other teachers through a similar process of engagement. Also, they will introduce their students—primarily in diverse urban schools—to experiences of mathematical exploration and to reflection on math-team discourse. The model of math teacher professional development and of student collaborative math learning centers on the production of significant math discourse.

Theoretical Framework: Math Cognition as Math Discourse

To mathematicians since Euclid, math represents the paradigm of creative intellectual activity. Its methods set the standard throughout Western civilization for rigorous thought, problem solving and argumentation. Many of us teach math in part to instill in students a sense of deductive reasoning. Yet, too many students—and even some math teachers—end up saying that they “hate math” and that “math is boring” or that they are “not good at math” (Boaler, 2008; Lockhart, 2009). They have somehow missed the intellectual math experience—and this may limit their lifelong interest in
Proposals for Research

science, engineering and technology. According to a recent “cognitive history” of the origin of deduction in Greek mathematics (Netz, 1999), the primordial math experience in 5th and 4th Century BC was based on the confluence of labeled geometric diagrams (shared visualizations) and a language of written mathematics (asynchronous collaborative discourse), which supported the rapid evolution of math cognition in a small community of math discourse around the Mediterranean, profoundly extending mathematics and Western thinking.

The vision behind our project is to foster communities of math discourse in networks of math teachers, in classrooms of K-12 math students and in online communities associated with the Math Forum. We want to leverage the potential of networked computers and dynamic math applications to catalyze groups of people exploring math and experiencing the intellectual excitement that Euclid’s colleagues felt—refining and testing emerging 21st Century media of collaborative math discourse and shared math visualization to support math discourse in both formal and informal settings and groupings. Those members of the project team who teach math teachers masters-level courses and professional-development workshops—and others—have found that many people teaching K-12 math have had little experience themselves participating in processes of mathematical exploration and discovery (Krause, 1986; Livingston, 1999; Silverman & Thompson, 2008). This project is designed to provide teachers with first-hand experiences and to mentor them in guiding their students to engage in rich math discourses that go beyond generating numeric answers to supply math reasoning and to draw conceptual connections (Briedenbach et al., 1992; Carlson, 1998; Carlson et al., 2002; Monk, 1992; Thompson, 1994).

The learning sciences have transformed our vision of education in the future (Sawyer, 2006; Stahl, Koschmann & Suthers, 2006). New theories of mathematical cognition (Bransford, Brown & Cocking, 1999; Brown & Campione, 1994; Greeno & Goldman, 1998; Hall & Stevens, 1995; Lakatos, 1976; Lemke, 1993; Livingston, 1999) and math education (Boaler, 2008; Cobb, Yackel & McClain, 2000; Lockhart, 2009; Moss & Beatty, 2006), in particular, stress collaborative knowledge building (Bereiter, 2002; Scardamalia & Bereiter, 1996; Schwarz, 1997), problem-based learning (Barrows, 1994; Koschmann, Glenn & Conlee, 1997), dialogicality (Wegerif, 2007), argumentation (Andriessen, Baker & Suthers, 2003), accountable talk (Michaels, O’Connor & Resnick, 2008), group cognition (Stahl, 2006) and engagement in math discourse (Sfard, 2008; Stahl, 2008). These approaches place the focus on problem solving, problem posing, exploration of alternative strategies, inter-animation of perspectives, verbal articulation, argumentation, deductive reasoning and heuristics as features of
significant math discourse (Maher, Powell & Uptegrove, 2010; Powell, Francisco & Maher, 2003; Powell & López, 1989).

To learn math is to participate in a mathematical discourse community (Lave & Wenger, 1991; Sfard, 2008; Vygotsky, 1930/1978) that includes people literate in and conversant with topics in mathematics beyond basic arithmetic. Learning to “speak math” is best done by sharing and discussing rich math experiences within a supportive math discourse community (Papert, 1980; van Aalst, 2009). By articulating thinking and learning in text, students make their cognition public and visible. This calls for a reorientation of the teaching profession to facilitate dialogical student practices as well as requiring content and resources to guide and support the student discourses. Teachers and students must learn to adopt, appreciate and take advantage of the visible nature of collaborative learning. The emphasis on text-based collaborative learning can be well supported by computers with appropriate computer-supported collaborative learning (CSCL) software, such as that prototyped in the Virtual Math Teams (VMT) Project (Stahl, 2009b).

Research Project Goal, Hypothesis and Components

Project Goal

To incrementally refine a research-based, classroom-tested model of computer-supported, resource-supported math education through shared visualizations and collaborative discourse by groups of mentored teachers and groups of their students—by designing, developing and testing: (i) a discourse-based model of math-teacher professional development and mentoring support; (ii) customized technology for computer support of shared math visualization and joint exploration; and (iii) adaptable, standards-based math-content teaching resources for middle-school and high-school students, guidelines for group collaboration and accountable talk, tools for reflection on discourse and networks of on-going mentoring relationships for math teachers.

Research Hypothesis

The project is based on an hypothesis, which it will test, concerning how to increase the quality and quantity of significant math discourse among math teachers and K-12 students:
Indicators of math learning (by groups of teachers and groups of their students)—such as group discussion of math content, problem posing/exploring/solving, explanation of math moves, visualization or investigation of multiple representations, and reflexive analysis of group math work—can be increased through (i) a math-discourse-based model of in-service teacher professional development supported by and integrated with use of (ii) a multi-user version of dynamic mathematics technology integrated in a rich online learning environment to support shared visualization and joint exploration of mathematical topics and (iii) mentoring relationships, collaboration and accountable talk guidelines, and curricular resources for online professional-development courses, K-12 classes and formal and informal online math communities.

This hypothesis is intended to guide iterative cycles of trial and analysis in design-based research (design, develop and test—not to prove efficacy and effectiveness). It will assess the effect of the combination of project components—because in such a socio-technical system the effect of introducing the technology is highly dependent upon the mentoring and the use of appropriate resources.

The hypothesis centers on measurements of group math discourse rather than on assessment of individual learning of math content—in accordance with the socio-cultural theory that effective individual math learning can be an indirect product of participation in group math discourse (Lave & Wenger, 1991; Sfard, 1998; 2008; Stahl, 2006; Vygotsky, 1930/1978). Vygotsky's notion of the zone of proximal development suggests that students may be able to engage in mathematical work within groups at a level that they will not be able to engage in for a couple years as individuals—and that such group work can be essential for the individual development in the long run (Vygotsky, 1930/1978, pp. 84-91). As a result, there is a need to assess the educational effectiveness of group interactions as such, beyond pre/post tests of the individuals. In addition, the striking finding within CSCL research of productive failure (Barron, 2003; Kapur & Kinzer, 2009; Patak et al., 2011; Schwartz, 1995) shows that there can be a paradoxical inverse relationship between measures of successful learning by small groups versus by the individual members of those groups because of group processes that reveal deep mathematical relationships but that do not lead immediately to high test scores of the individuals. For these reasons, the project evaluates its goal in terms of the quantity and quality of the math discourse that takes place during the small-group problem-solving interactions, looking for hypothesized increases for groups as they
participate and in successive project years as the model, technology and resources are iteratively developed.

(i) Model of Math Education

The proposed project will design, develop and test a model of math education through collaborative math problem proposing/exploring/solving, by involving in-service teachers in first-hand mathematical experiences and helping them to reflect on their own learning experiences. Then they will try out the model with their students, while receiving mentoring and support from the project. The collaborative model of math education stresses math discourse. In this project, groups of teachers and groups of students will do math problem solving collaboratively and then reflect on the logs of their discourse to identify key moves. We propose using teachers’ and their students’ original mathematical conversations as “didactic objects” (Thompson, 2002) designed to support “decentering” (Wolvin & Coakley, 1993) and “collective reflection” (Cobb et al., 1997) on particular aspects of their math discussion. The discourse-centered model of math education will structure learners in small teams and will provide mentoring to guide the team’s mathematical exploration, discourse and learning. Math Forum staff and other project team members will provide initial mentoring to the first cohorts of teachers, who will in turn mentor subsequent cohorts of teachers as well as students in their own classes. A permanent support network will be established to provide sustainability of project accomplishments. The teachers who are trained in this project will be encouraged—initially by paying them—to participate in teacher networks, including national and international networks of teachers, supporting broadening dissemination of the discourse model of math education.

(ii) Online Math Collaborative Learning Environment

The proposed project will design, develop and test two forms of technology to support math learning with collaborative and interactive tools for cyberlearning: (a) computer-supported collaborative learning (CSCL) software and (b) dynamic mathematics (software that allows users to manipulate geometric diagrams, equations, etc.). (a) CSCL provides virtual learning environments in which teams of students can interact synchronously and asynchronously to build knowledge together. This student-centered approach has many advantages, including increased motivation, sharing of skills, engaging in significant discourse and practicing teamwork. This project will adapt and extend the Virtual Math
Teams (VMT) environment already prototyped and tested by the PIs (Stahl, 2009b). (b) Dynamic math (such as Geometer’s Sketchpad, Mathematica, Cabri or GeoGebra) has already profoundly impacted math education (Goldenberg, 1995; Hoyles & Noss, 1994; King & Schattschneider, 1997; Laborde, 1998; Myers, 2009; Scher, 2002), with Geometer’s Sketchpad and GeoGebra used in many US classrooms and globally. Yet, research on math education has not analyzed how students use dynamic math tools in sufficient detail (compare Çakir, Zemel & Stahl, 2009; Stahl, 2009b). GeoGebra (http://www.geogebra.org) is an open-source system for dynamic geometry, algebra and beginning calculus—including trigonometry, conics, matrices, graphing and Euclidean constructions. It offers multiple representations of objects in its graphics, algebra and spreadsheet views that are all dynamically linked, making GeoGebra a particularly flexible tool for exploration. Working with the developers of GeoGebra, this project will provide the first multi-user version of dynamic math, so that teacher teams and student teams can explore math collaboratively; it will integrate this into the larger VMT virtual collaborative-learning environment with text chat and wiki to support persistent discourses about math—that can be shared, reflected on and researched.1

1 For a demo of the prototype system, go to http://vmt.mathforum.org/VMTLobby. Log in as “guest” with password “guest”. The Lobby should open showing the List of All Rooms. Select Project “VMT Research”. Click on "Apply filters". Open “Geometry”. Open “Polygons”. Click on "GeoGebra Demo Room" Eventually a JavaWebStart chat room should open. Explore its different tabs and functions.
Proposals for Research

Figure 1. A demo (not real student interaction data) GeoGebra construction created and discussed collaboratively in a proof-of-concept multi-user prototype of the project’s learning environment, based on the VMT system. The VMT system includes (not shown here): a Lobby with social networking and tools for teachers, integration with a wiki, and Web browsers.

(iii) Curricular Resources

The proposed project will design, develop and test resources to support teachers and students in their interactive explorations of rich math problems (e.g., open-ended problems with multiple possible solution approaches and many potential extensions to explore). Three kinds of resources are: (1) Curriculum packages in domains of K-12 math, building on existing NSF-funded and community-based sources (see http://dynamicgeometry.com, http://keypress.com/x5582.xml and http://geogebra.org/en/wiki). The curriculum will be based largely on classroom-tested problems using dynamic-math software and integrated with popular math textbooks (e.g., Everyday Mathematics, Investigations in Number, Data and Space, Mathematics in Context, Connected Mathematics, Interactive Mathematics Program, Core-Plus Mathematics, Simms Integrated Mathematics and textbooks from McDougal Littell or Glencoe), but adjusted by experienced Math Forum staff for collaborative online usage. It will be aligned with the recommendations of the Common Core
State Standards for Mathematics and the new NCTM volumes, Focus in High School Mathematics: Reasoning and Sense Making in Algebra/Geometry. Teachers will be mentored in adapting the content of their local curriculum to collaborative online student exploration, whether using GeoGebra or not. (2) Guidelines, suggestions and examples for collaborative learning, knowledge building and math exploration will be published. This will feature “accountable talk” guidelines for math discourse. (3) Training resources in understanding online math discourse will be developed to help teachers and students identify examples of productive inquiry moves, etc., to foster reflection on logs of their math discourses. These broad categories of resources will encapsulate the expertise of the project team in problem design, collaboration mentoring and discourse analysis, producing documents that can be used by a gradually growing community of math teachers and students. The content of these resources will build on experience at the Math Forum, the VMT Project, the teacher professional-development programs at Drexel and Rutgers and the related research literature. The content will be elaborated, tested, evaluated and refined—and then published as project deliverables.

Results from Prior NSF Support

The proposed project grows out of the successful Virtual Math Teams (VMT) Project. This is a several-year NSF project (awards DUE-0333493, IERI-0325447, SBE-0518477, DRL-0723580) that developed an open-source virtual learning environment for math students. The system integrated a social-networking portal, synchronous text chat, a shared whiteboard, an asynchronous wiki, a referencing tool, mathML expressions and a web browser. Student actions and chat postings are automatically logged to be replayed for analysis. Over a thousand student-hours of piloted usage were logged. A qualitative micro-analytic approach to interaction analysis was developed based on ethnomethodologically inspired conversation analysis (Garfinkel, 1967; Sacks, 1962/1995; Stahl, 2009a; 2009c; Zemel, Çakir & Stahl, 2009). A large number of publications have appeared from the project (see http://GerryStahl.net/vmt/pubs.html), including 2 books (Stahl, 2006; 2009b) and 6 doctoral dissertations (Çakir, 2009; Litz, 2007; Mühlpfordt, 2008; Sarmiento-Klapper, 2009; Wee, 2009; Zhou, 2010).

The VMT Project pioneered the study of online collaborative math discourse—both its nature and modes of computer support for it. The 28 studies in (Stahl, 2009b) present some of the most important of the 169 publications related to the project. They include a number of dissertation-
level case studies of interactions in the VMT environment by middle-school, high-school and junior-college students, which analyze: how math problem solving can be effectively conducted collaboratively among students who have never met face-to-face; how the structure of text chat interaction differs from spoken conversation; how the media of graphical diagrams, textual narratives and symbolic representations can be intimately interwoven to build deep math understanding; how deictic referencing is important to establishing shared understanding; how students co-construct a joint problem space; how collaborative meaning making and knowledge building are accomplished in detail; how online math discourse can be supported by a software environment that integrates synchronous and asynchronous media with specialized math tools; and how a methodology based on interaction analysis can be used for a science of group cognition.

The VMT Project was structured as design-based research, with the technology, research and theory co-evolving through dozens of iterations. The VMT Project demonstrated both the practicality of the proposed project and the need for it. While the VMT Project prototyped a rich cyber-learning environment and studied student interaction, it did not develop the range of supports that we know are needed for classroom use: robust software, problem sets, guidelines, etc. Furthermore, it did not include a dynamic-math component. The VMT Project provides a solid starting point for the proposed project and documents the need for further technological development, enhanced support for dynamic math, curricular models and training of in-service teachers. The design, development and testing of these logical next steps are needed to enable a powerful and innovative form of math education to be offered in a practical form to K-12 schools through education schools and to the public through the Math Forum.

Prior NSF support of the Math Forum has developed a successful approach to online mentoring of math teachers and their students. Since 1993, the Math Forum has mentored over 100,000 students, conducting hundreds of workshops, summer institutes and school-improvement contracts. Recently, it has successfully completed the Virtual Fieldwork, Online Mentoring, and Teacher Workshop Model projects (NSF DUE-0717732, DUE-0127516 and DUE-0532796). Mixed-methods studies of these have shown the surprising result that the online mentoring of K-12 pre- and in-service teachers had a more positive effect for teachers with low math self-efficacy (Renninger et al., in press). This is due to the non-linear and flexible format of online discussion—suggesting that online collaboration may well help at-risk math students at least as much as those with higher math self-efficacy. Math Forum approaches are making inroads with a population of people who most would think will not change
(Renninger et al., 2010). In the proposed project, Math Forum workshops for teachers will complement and feed teachers into the courses at Rutgers and Drexel. The workshops will also train mentors and seed the on-going teacher network.

**Research and Development Design**

The proposed project adopts an iterative design-based-research approach to design, develop and test innovative curriculum materials, technologies, teaching methods and models for teacher in-service professional development and K-12 student instruction. The project develops a socio-technical educational model that evolves and integrates a number of mutually supportive components, each of which has previously been explored in a preliminary way by one or more of the PIs. However, the components have not previously been integrated into a scalable model of math education. The proposed project brings together the PIs, other necessary senior staff and advisors with the resources to begin to systematically test, refine, validate and disseminate the integrated model. There are several areas of work:

(a) A **model of math education as computer-supported math discourse**. The model incorporates: (b) innovative technology for collaborative math discourse, (c) support for shared mathematical visualizations and (d) curricular materials to stimulate and guide math discourse. The model includes three successive project targets: (e) in-service teacher professional development, (f) middle- and high-school math education and (g) broader virtual math-discourse communities.
Figures 2 and 3. Images of actual student online collaborative work on patterns. In Figure 2, a student points from a chat message to a smallest hexagon pattern composed of 6 triangles illustrating VMT's unique integration of chat and whiteboard with its deictic reference tool. Figure 3 shows the Replayer tool interface across the bottom.

(b) **Innovative technology for collaborative math discourse.** The VMT Project developed a research prototype of a custom, open-source virtual learning environment that integrates synchronous (text chat, shared whiteboard, dynamic math exploration, shared web-browser) and asynchronous (a community wiki, a social-networking portal) media to support math visualization and collaborative discourse by virtual math teams. This prototype was adequate for extensive testing in multiple iterations, as well as limited use by select teachers in their classrooms as part of research trials. As part of the proposed project, we will implement, test and refine new interfaces for teachers, mentors and administrators. These will allow teachers to register a number of students at once, set up multiple copies of interaction rooms for multiple small groups of students, monitor activity in rooms, respond to problem behavior online and review reports of student activity. New functionality will also make it easier for students to document their online work (e.g., in the project wiki or in Word documents, Excel spreadsheets and PowerPoint slides) with log excerpts and images of constructions. Support for researchers will facilitate researchers in the project as well as colleagues outside the project to easily replay sessions of student interaction.

The VMT Project was widely recognized as an important example of synchronous support for online collaboration and was studied by several international researchers ([GerryStahl.net/vmt/pubs.html](http://GerryStahl.net/vmt/pubs.html)); it is expected that the proposed project will be of even more interest, particularly within the math education research community. The VMT Replayer allows complete replay of a user session, including all actions and system notices, as though the session was digitally video-recorded. The researcher’s view is guaranteed to be identical to the user’s view since it is generated from the same data as sent to a client computer. The log information will be made available in convenient textual formats for student reflection and reporting as well as for researcher analysis. New functionality to be explored includes automated feedback agents and displays, increased integration so math objects can be moved easily from the synchronous tabs (chat, whiteboard, summary, GeoGebra, web browsers) to asynchronous components (wiki pages, email, documents), as well as refinement of the interface. The system will be released as open source on SourceForge so that others can deploy it.
on their own servers or extend the software to meet their own educational needs. The Math Forum will maintain the system as a permanent service, so that users can easily create topics for chat rooms and invite other users to collaborate.

(c) Support for shared mathematical visualizations. The project will port GeoGebra—a comprehensive and well established application for dynamic-math exploration—to the virtual learning environment described above. It will make the application fully multi-user. It will integrate the application in a tab of the environment (see Figure 1 above). As previously described, GeoGebra is a particularly appropriate dynamic-math application for this project because its source code is freely available as open source, there is an active international development community to support on-going development, the lead developer and the founder are committed to consult on this project, the application supports a wide range of math from algebra and geometry construction to calculus and 3-D, GeoGebra has won international prizes, it has been translated into about 50 languages and it has received on-going NSF support. Like all other dynamic-math applications, GeoGebra currently exists only as a single-user application. While users can send their static constructions to each other, display screen images, or awkwardly include a view of the GeoGebra application within other environments through screen sharing (e.g., in Blackboard, Moodle, Elluminate, etc.), only one person can dynamically manipulate the construction. Our port converted GeoGebra to a client-server architecture, allowing multiple distributed users to manipulate constructions and to all observe everyone’s actions in real time. Every action in the GeoGebra tab will be immediately broadcast by the server to all collaborating clients (and logged in detail for replay and research). We have been exploring turn-taking mechanisms (see Figure 4) to avoid conflicts in the construction and modification of GeoGebra drawings; although it is important in synchronous chat to allow multiple users to type simultaneously, we have found that it is natural for a group to allow one member at a time to change a graphical construction and for group members to take turns editing and rearranging.
Incorporation of GeoGebra in the VMT environment framework allows users to engage in text chat while manipulating the construction. Importantly, users can graphically point from a chat posting to an area of the construction that they want to index (see Figure 2)—an important support for math discourse that is unique to VMT. They can also scroll back and forth through the history of the GeoGebra construction, animating its evolution—a powerful way to explore many mathematical relationships. In addition, a complete record of the collaborative construction is available to the participants, their teachers and project researchers, allowing them all to analyze and reflect upon the complete interaction, including the construction actions synchronized with the chat. We have already completed a prototype port of GeoGebra to VMT in order to confirm its feasibility. It provides an exciting collaborative experience. The port now needs to be engineered in a robust way, incorporating all of the GeoGebra functionality (including import and export compatible with standard GeoGebra and Geometer’s Sketchpad to facilitate sharing of constructions, and a full menu system to support learning by new users). In Year II of the project, we will incorporate the extended GeoGebra 4.0 functionality that will be released by then, including support for inequalities and CAS (computer algebra system like Mathematica, Maple, or the TI-Nspire CAS calculator). The project will produce a refined and tested multi-user version of GeoGebra and will release it as open source.

(d) Curricular materials to stimulate and guide math discourse. Problem-based learning (PBL) materials in areas of mathematics like algebra, combinatorics and geometry will be adapted from existing high quality curricula and piloted. These materials will define challenging math problems for collaborative online group exploration and help teachers to tune them to local student capabilities. The materials will allow students to explore rich but accessible problems taken from topic domains covered in their textbooks and in the Common Core State Standards. The PBL approach involves mentors who are trained to guide student exploration and to steer collaborative student groups to address their joint learning issues (Barrows, 1994; Hmelo-Silver, 2004; Hmelo-Silver & Barrows, 2008; Koschmann, Glenn & Conlee, 2000). Project team members and others have developed some model math problems (Krause, 1986; Math Forum & Wolk-Stanley, 2003a; 2003b; 2004a; 2004b; Powell, Lai & O’Hara, 2009). The Math Forum has years of Problems-of-the-Week in several areas of school mathematics, which can be adapted to online group collaboration. Much curriculum has been developed with NSF funding for dynamic-math applications like GeoGebra and Geometer’s Sketchpad, including lessons tied to state standards and intended to support popular textbooks through
student hands-on exploration. The project will facilitate classroom teacher use of such resources in this new learning context. The team has already prototyped a series of problems that consecutively explore issues of combinatorics; along with the problems, a teachers’ guide contains concrete suggestions on how to adapt the problems for different kinds of student teams (Powell, Lai & O’Hara, 2009). The problems in this document were tested in the VMT Project and in high-school classrooms of teachers studying at Rutgers. Sets of problems correlated to textbooks and to the Common Core State Standards will be compiled, some taking advantage of GeoGebra. Additional resources will be developed to train teachers and students in mentoring techniques, in collaboration skills and in math-discourse skills. All these resources will be tested and produced in publically available online documents as project deliverables. These and other math problems will be incorporated in the VMT Lobby’s library of Topics, to be available to students in home-schooling and informal-learning situations.

(e) In-service teacher professional development. To effectively change education in schools, teachers must be prepared to understand and to learn how to model use of the innovative technologies and pedagogies. Practicing teachers rarely find time to engage in learning processes capable of transforming their teaching practice and they seldom are able to introduce major new approaches in their highly constrained curricula. This project therefore involves in-service teachers when they have scheduled time to pursue masters-level professional-development courses. It starts by involving them during their regular courses (taken online) in online collaborative problem solving using the project’s software technology and curricular approach—(a), (b) and (c) above. Later course work involves them trying out what they have learned back in their own classrooms, within the context of their current curriculum; the project provides mentoring and resources to support this effort.

Both Drexel and Rutgers-Newark offer masters-level teacher-professional-development programs and courses in math education in online modes. The fact that these teachers will already be studying together online creates an ideal setting for the use of an online learning environment with dynamic-math support. These graduate programs have been designed, taught and directed by project co-PIs Silverman and Powell. The proposed project will allow these programs to develop, test and adopt the educational model of computer-supported math discourse. This model will be pioneered at these two schools of education, providing a collaborative interaction that will produce a more generalized result than would development at a single institution. It will also permit extended utilization of the online medium by,
for instance, having teachers from both institutions working together on
math topics in small groups and having them mentor teachers from each
other’s institution. In the later years of the project, this model will be
disseminated to other schools of education, partially through Advisory
Committee members. The Math Forum has effectively implemented a
similar model, incorporating its Online Mentoring Project modules into
teacher education programs around the country.

The initial plan at *Drexel University* is to build on the existing MS in
Mathematics Learning and Teaching (MS-MLT) program, which is already
exclusively offered online. This program in math education was originally
developed by co-PI Silverman and is taught primarily by him and Math
Forum staff. For the first cohort of students under this project, Drexel will
offer MTED775, “Special Topics: Supporting Math Learning through
Computer-Supported Collaborative Discourse.” This course will be one
required math-education elective for MS-MLT students and an elective for
other professional-development students. Then two new education courses
will be developed to make this model a part of the regular course offerings
of the School of Education: MTED 651 (which will focus on teachers
personally engaging in computer-supported, resource-supported
collaborative discourse and reflection on both their activity and their
learning) and MTED 652 (which will focus on supporting teachers to
incorporate computer-supported, resource-supported collaborative discourse
in their classes). MTED 652 will include resource development for teachers' 
classroom implementations. Each of these courses—which have been
approved at Drexel pending funding of this project—will carry 3 quarter-
credits.

The initial plan at *Rutgers-Newark* is to engage two cohorts each year
of practicing teachers in a revised version of the online course in
“Mathematics and Instructional Technology” taught by co-PI Powell. The
goals of the course are three-fold: (1) to familiarize in-service teachers with
the mathematical problem-solving and problem-posing activities of the
online problem-exploration units in which their students will engage; (2) to
depth in-service teachers’ thinking about the effects of the collaborative
environment on their own and their students’ thinking about mathematics
(math objects, relations among objects and dynamics among relations), math
reasoning and problem-solving heuristics; (3) to focus in-service teachers’
instructional attention on understanding and facilitating students’ discourse
in mathematics. To accomplish these goals, the course will engage in-
service teachers in a sequence of tasks, beginning with familiarizing them
with the project online environment through involving them in mathematical
activities using it, then engaging them in reviewing their session logs and
finally having them plan how they will implement the model in their teaching.

**f) Middle- and high-school math education.** The in-service teachers will introduce the technology and curricular resources that they used in their university classes into the classes they teach, often mixing students from different schools or cities in online teams to take advantage of being part of an online discourse community and to motivate the use of online media by students in face-to-face classrooms. The teachers will take the logs of their students’ interactions back to their professional-development sessions for on-going group analysis. They also will engage their students in reflection on their own logs, discussing how the math discourse surfaces mathematical insights and conceptual connections.

The curricular resources adapted by the project are designed to support classroom math activities by enhancing and reinforcing the core objectives covered in textbook readings and instructor-led activities. Resources include adaptation options and guidelines to help teachers tune problem sets to complement their core activities. For instance, the research-based textbooks, *Mathematics in Context* and *Discovering Geometry*, which are used in the Philadelphia public school system, stress student investigation in order to construct conceptual understanding of key math concepts and the *Common Core State Standards for Mathematics* recommend that “students consider the available tools [such as] dynamic geometry software…to explore and deepen their understanding of concepts” (p.7). The project model builds on this approach, providing opportunities for students to explore and discuss topics online with peers from their own or other schools. The model provides: tools for dynamic, multi-user, graphical exploration; visual and numeric feedback on quantitative and qualitative changes during exploration; and a record of the exploration and accompanying discourse, which students can replay, reflect on and incorporate in reports—e.g., pasting log excerpts or screen images in their documents.

Reflection on interaction logs by teachers and students primarily involves trying to follow the problem-solving path of participants and to notice critical collaboration moves. They will be encouraged to look for examples of accountability to the group, to standards of math reasoning and to the characteristics of their math objects. They will look for instances where someone poses a productive inquiry that initiates effective group exploration—or where the group fails to come up with a useful proposal or fails to take up a proffered proposal. Examples will be culled and shared on the project wiki.
Although many project activities center on teacher professional development, the ultimate goal is to increase the quality and quantity of both teacher and student mathematical discourse. Therefore, teacher professional development will be oriented to improving the math discourse of their students. While the primary indicator of project success will be the identification of desirable mathematical discourse moves during problem solving by teachers and students, the project will also be concerned with changing student conceptions of math. It will survey a sample of teachers and students before and after their involvement in the project to compare self-reports of attitudes about math and about approaches to math instruction. In addition, some teachers and students will be asked as a final part of their course work to compose a brief reflection paper on their learning experience.

Most of the in-service teachers in the project come from the Philadelphia, Camden, Newark, New Brunswick and New York City areas. Thus, many of the classrooms that will be involved in the program are inner-city K-12 schools with high proportions of educationally at-risk and economically disadvantaged students; others are from near-by suburban and private schools with contrasting student populations. The project educational model will therefore be tested in diverse, real-world settings.

Because teacher and student work on math problems will all take place in the online software environment, complete detailed logs will be available to the project staff, as well as to the students and teachers themselves. The logs can be reviewed and studied in detail with the Replayer software, as well as with various formats of log printouts. This will not only facilitate reflection by students and teachers on their own work, but also permit the documentation of interesting cases for teacher instruction and detailed analysis for project evaluation. The project will compile a portfolio of instructive case studies.

(g) Broader math-discourse communities. Once teachers studying at Drexel or Rutgers and their students become involved in online collaborative dynamic geometry and math discourse, teams will be set up that involve students from online schools, home-schooling networks or the Math Forum virtual community. This will yield data for generalizing project findings as well as stimulate the spontaneous generation of self-organizing communities of math discourse. This will primarily take place through contacts and presentations by project staff and the teachers who have been trained, as well as through the Math Forum and its large user community (3 million visits/month. The project technology and resources will be made publically available as an integral part of the Math Forum services in Years
IV and V of the project. The VMT software environment is designed to support the viral spread of user communities across the Internet; the proposed project is intended to form a critical mass of users and topics to catalyze that process. The model of computer-supported math discourse will become institutionalized at Drexel and Rutgers, will be taken to other schools of education through Advisory Committee members and personal contacts of project staff, through Math Forum outreach, through the extensive active GeoGebra user community and through presentations at educational conferences and in related journals.

(h) Group cognition theory. When small groups engage in collaborative problem posing, exploring and solving, they can accomplish cognitive tasks interactively or transactively as a group. The project will analyze logs of student math work, shared visualizations and reflective discourse, using conversational analysis and statistical methods to study how students build on each other’s utterances, constructions and actions to accomplish mathematical cognition. Building on past work on group cognition (Çakir, Zemel & Stahl, 2009; Koschmann, Stahl & Zemel, 2009; Stahl, 2006; 2010a; 2010b), this will provide a contribution to theory of situated and distributed cognition. In particular, analysis of the use of GeoGebra in a fully logged multi-user online environment with guidance in math discourse moves will pioneer in the development of theory of cognition in groups using dynamic-math tools, providing insight into math learning generally. Case studies and other findings with theoretical implications will be published.
<table>
<thead>
<tr>
<th>Resources</th>
<th>full logging teacher admin &amp; monitoring supports</th>
<th>statistical analysis</th>
<th>participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test teacher and student resources</td>
<td>Evaluate use of resources</td>
<td>Publish resources</td>
<td>Disseminate resources</td>
</tr>
</tbody>
</table>

| Curriculum materials                                                     | Review existing materials for GeoGebra and Geometer's Sketchpad | Compile problem sets aligned with standards and textbooks | Evaluate use of materials | Publish materials in formats for teachers, home schooling, distributed schooling | Disseminate materials |

| Teachers                                                                 | Pilot model with 10 teachers in Drexel and Rutgers courses and 20 teachers in Math Forum workshops | Implement model with 35 teachers in Drexel and Rutgers courses and 40 teachers in Math Forum workshops | Evaluate model with 50 teachers in Drexel and Rutgers courses and 40 teachers in Math Forum workshops | Continue training with 60 teachers in Drexel and Rutgers courses and 40 teachers in Math Forum workshops | Evaluate teacher training in Drexel and Rutgers courses; continue training 40 teachers in Math Forum workshops |

| Students                                                                 | Pilot with 25 students | Involve 750 students of teachers in courses and workshops; log series of sessions by student small groups | Involve 750 students of teachers in courses and workshops; log series of sessions by student small groups | Involve 750 students of teachers in courses and workshops; log series of sessions by student small groups | Evaluate changes in significant math discourse of student groups over time: within group and across cohorts |
Proposals for Research

<table>
<thead>
<tr>
<th>Mentoring</th>
<th>Project Phases, Milestones, Deliverables</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mentoring</strong></td>
<td>Prepare mentoring materials based on previous Math Forum mentoring projects</td>
</tr>
<tr>
<td></td>
<td>Pilot mentoring of teachers with 2 outstanding teachers</td>
</tr>
<tr>
<td></td>
<td>Increase to 5 teacher mentors</td>
</tr>
<tr>
<td></td>
<td>Increase to 10 teacher mentors</td>
</tr>
<tr>
<td></td>
<td>Increase to 15 teacher mentors</td>
</tr>
<tr>
<td><strong>Theory</strong></td>
<td>Validate coding scheme</td>
</tr>
<tr>
<td></td>
<td>Analyze discourse moves in logs</td>
</tr>
<tr>
<td></td>
<td>Conduct in-depth case studies and interviews</td>
</tr>
<tr>
<td></td>
<td>Compile best practices case studies</td>
</tr>
<tr>
<td></td>
<td>Develop theory of math group discourse</td>
</tr>
</tbody>
</table>

**Evaluation**

Formative evaluation is a constant process built into the design of the project. As a design-based research effort, the over-riding research hypothesis listed at the start of this project description will be addressed by designing and exploring an iteratively refined solution—and by documenting its impact on the quantity and quality of math discourse by teachers and students. The interlocking components of the project will be reviewed at weekly project team meetings. Team meetings will include interaction analysis data sessions (Jordan & Henderson, 1995; Stahl, 2010a), in which the group collaboratively discusses new data from logs of teachers or students—and makes design decisions for refining the co-evolving components. The project team will discuss what seems to be working and what does not. It will decide what to modify for the next iteration. The project is complex, with many dependencies among its components and many shifting contextualities. A flexible approach like design-based research is needed to respond to a continuous formative evaluation and ongoing project modification.

The explicit evaluation effort will include semi-annual formative-assessment reports documenting: (a) project progress, (b) improvements in project outcomes and (c) plans for the next half year. The external Advisory Committee (AC) will review, discuss and respond to each report. The AC will meet annually to discuss project progress with the project team. The AC has expertise in mathematics education, research evaluation, teacher training, problem-based learning, conversation analysis, CSCL and virtual...
as discussed above, the research hypothesis focuses on the quantity and quality of math discourse at the group unit of analysis. Theories of the zone of proximal development, productive failure and group cognition argue that learning-related processes and phenomena at the group level may be different from those at the individual level. Other research has documented the efficacy of dynamic-math visualization tools for individual learning; for instance, a study of geometry students in eleven Florida schools revealed a significant difference in the FCAT mathematics scores of students who were taught geometry using Geometer’s Sketchpad compared to those who used the traditional method—regardless of differences based on SES or gender (Myers, 2009). The proposed project has a different focus. The PI and colleagues have developed coding schemes and analysis approaches oriented to the group unit of analysis based on conversation analysis of adjacency pairs and longer sequences (Sacks, 1962/1995; Schegloff, 2007; Stahl, 2009b, Chs. 20, 22, 23, 26; 2011b; Stahl et al., 2011). This approach serves both quantitative and qualitative analysis, by simultaneously specifying the structure of meaningful discourse moves and providing countable categories of group interaction units, in order to document changes over time—comparing discourse characteristics in selected time slices within teams or across cohorts.

The project will automatically produce raw data in the form of log files of participant online interactions. The log files are anonymous, but allow tracking of individual users through consistent login handles. The VMT environment is instrumented to capture all user actions in the chat and whiteboard—this will be extended to multi-user GeoGebra. A database of all sessions is automatically maintained and provides spreadsheet logs in handy formats and Replayer files. Software tools will be used for automated and manual log analysis of discourse measures and their evolution during training. While low-level group processes (e.g., number, length and rate of chat postings and drawing actions in different time slices) can be tracked
automatically and analyzed statistically, higher-level math-discourse processes have to be interpreted manually. The PI has on-going, NSF-supported collaborations with Carolyn Rosé of Carnegie-Mellon University’s intelligent tutoring group, exploring software agents in the VMT environment to provide student guidance and also investigating computer support for coding discourse moves in text chat, to aid and supplement manual analysis. Raw and coded logs will be maintained in a database to facilitate analysis of changes over time for groups across sessions and across successive cohorts of participants.

Quantitative analysis—based largely on the coding of discourse moves in teacher and student VMT logs—will track changes in key measures of significant math discourse. The project hypothesis will be operationalized as predicting an increase in specific measures as a given group works in the VMT environment during time slices across an academic term. Logs of the following groups involved in the project will be evaluated: (a) in-service teachers participating in Math Forum workshops, (b) teachers working together as part of teacher professional development course work, (c) students guided by their teachers, (d) students working with other students as part of school classes and (e) students interacting with others informally at other schools or globally.

Discourse will be coded and measured along the following dimensions: (1) volume of discourse and level of participation, (2) percentage of on-task math discourse, (3) use of representations, (4) integration of chat and drawing, (5) use of accountable talk moves, (6) adoption of socio-mathematical norms and practices, (7) speaking meaningfully with explanation and argumentation, (8) involvement in posing, exploring and solving problems and (9) additional dimensions to be developed based on project experience. The theory of math learning through participation in math discourse (Sfard, 2008) specifies important mathematical discourse moves, such as encapsulation, reification, saming, routines, deeds, explorations and rituals. The theory of accountable talk (Michaels, O’Connor & Resnick, 2008; Resnick, 1988) specifies discourse moves that promote accountability to the group, to standards of math reasoning and to the characteristics of the math objects. Speaking meaningfully in math discourse “implies that responses are conceptually based, conclusions are supported by a mathematical argument and explanations include reference to the quantities in the problem context [as opposed to a focus on merely describing the procedures and calculations used to determine the answer” (Clark, Moore & Carlson, 2008, p.298). Socio-mathematical norms include what counts as an acceptable, a justifiable, an easy, a clear, a different, an efficient, an elegant and a sophisticated explanation (Yackel, 1995; Yackel
Proposals for Research

& Cobb, 1996). Mathematical practices emerge from interaction, are taken up by participants and are applied repeatedly (Medina, Suthers & Vatrapu, 2009; Stahl, 2011a). These dimensions of significant math discourse are associated with typical sentences and discourse moves that can be identified by coders. A coding scheme will be validated with acceptable inter-rater reliability, as in (Stahl, 2009b, Chs. 22, 23; 2011b).

Detailed interaction analysis of selected cases will show how the math discourse actually evolves. Quantitative analysis can establish the statistical significance of changes in learning outcomes, but it generally does not provide much insight into the mechanisms of change; these mechanisms will become visible in detailed case studies in which the specifics of the interactions can be studied. By combining quantitative and qualitative analysis of discourse transformations, the project evaluation will determine how the online interaction involves engagement in significant mathematical discourse. This will help researchers to determine what to try in subsequent cycles of research and will allow evaluators to judge project progress.

**Summative evaluation** will assess the degree to which the discourse of teams of teachers and of students reveals—through the quantity and quality of their math discourse—increased understanding and improved practice of mathematics. It will make sure that project products (software, mentoring guides, problem sets, masters courses, analysis tools, best practices case library and analyses of case studies from the data corpus) have been produced and made publicly available. It will assess the effectiveness of these products based on the analyses of their use by teachers and students as logged in the data corpus, using quantity and quality of the facilitated math discourse as a measure of success.

In addition to the quantitative and qualitative analysis of changes in significant mathematical discourse by groups of teachers and students involved in the project, there will be ethnographic observations of participants. The observations—including pre/post surveys, open-ended interviews and reflection reports—will be primarily conducted by co-PI Khoo and External Evaluator Blanc, both trained cultural anthropologists. The goal of these observations will be to establish—as much as possible from the user perspective of the project participants—the effectiveness of project interventions (the pedagogical model, the technology, the resources). Interviews with students and teachers will explore their changed attitudes toward mathematics and their insights into the nature of mathematical reasoning. This will be triangulated with the analysis of the math discourse of the same participants in specific time slices. Ethnographic observation of teachers will additionally explore to what extent they have come to feel that
teaching math-discourse skills is key to fostering student math learning; to what extent they try to use the project model, technology and resources in their regular teaching; to what extent they intend to stay involved in support networks. The summative evaluation will report on these issues as well as the timely accomplishment of project tasks, training levels, dissemination efforts and project deliverables.

**Dissemination**

The primary avenues of dissemination will be: (a) through the Math Forum, (b) through Schools of Education, (c) through teacher professional associations, (d) through GeoGebra and dynamic math user communities and (e) through virtual learning communities, including home schooling and online schools.

(a) By the end of the project, the technology and the resources developed through the project will be publicly available as services of the Math Forum. The Math Forum has been the premier online resource for mathematics teaching and learning for over 16 years. It has three million visits to the site each month; its digital library contains over a million web pages, mostly user generated (as a forerunner of the Web 2.0 philosophy). Public services (which typically started from NSF-supported research projects) have been made sustainable through support from Drexel University, fee-for-service programs and teacher training contracts. The Problem-of-the-Week (PoW) is the Math Forum’s core service and is subscribed to by many school districts. It is primarily oriented toward problem solving of challenging math problems by *individual* students. The result of the proposed project would be to extend this service with open-ended math problems for *groups* of students to explore collaboratively online. Teachers using the PoW service would be encouraged to involve their students in the new service, initially interacting with classmates, but eventually joining cross-school, national and international virtual math teams. Math Forum services typically support both formal and informal mathematics learning by teachers and students (Renninger & Farra, 2003; Renninger & Shumar, 2002b; 2004; Shumar & Renninger, 2002).

(b) Several of the co-PIs and Advisory Committee members (e.g., Powell, Silverman, Derry, Hmelo-Silver, Hoadley, Koschmann, McClain, Renninger and Sfard) teach at schools of education across the country—and are in contact with math educators at many more. The project accomplishments will influence the teacher professional-development programs in these centers. Teachers who are involved in the teacher professional-development components of this project will also spread
project findings as early adopters at their graduate programs and K-12 schools. Ready access to project resources, models and technology at the Math Forum will facilitate general dissemination of innovative math education—including through the popular teacher discussion forums on the Math Forum website—to additional teacher professional development programs.

(c) The PIs and Math Forum are active in NCTM, AERA, PME, and PMENA and will present project findings at the annual conference for teachers of mathematics. Additionally, project researchers are prominent in the learning science communities around the ICLS, CSCL and other academic conferences and publish prolifically in academic and practitioner journals, books and conferences.

(d) Because it provides the first multi-user version of a dynamic-mathematics application, the project will be well known within the worldwide communities of GeoGebra and Geometer’s Sketchpad users. The project technology will all be available as open source, so that other researchers and developers can build on it, modify it and install versions on their own servers. (The project technology is built on VMT and GeoGebra, both already available as open source at SourceForge.) Teachers, trainers and researchers who do not have the technical expertise to do this, can simply use the environment that is on the Math Forum servers; they can develop their own curriculum for it and can readily access detailed user logs from it. Features for administration of chat rooms will be built in to support local administration.

(e) For the sake of sustainability beyond the proposed project and to support further scale-up, it is important to establish an on-going network of teachers in the form of self-organizing communities (Renninger & Shumar, 2002a). As discussed above, this will begin with mentoring relationships between cohorts of teachers going through the project professional development. The mentoring relationship will grow into a mutual support network, in which teachers from the programs at both Drexel and Rutgers will share questions, case studies, best practices, curriculum, etc. Later in the project, this growing local network will connect with national and international teacher networks, such as Tapped-In (http://tappedin.org), the Knowledge Building Teacher Network (Chan, van Aalst & Law, 2009) and the Institute for Knowledge Innovation and Technology (http://ikit.org). These networks will disseminate use of the project services widely. We are aware of the issues in trying to build sustainable virtual learning communities (Barab, Kling & Gray, 2004) and will use an iterative
approach. In addition, dissemination efforts will target organizations, consortia and networks of home schooling and of online schools.

Expertise

The proposed project brings together an interdisciplinary team of researchers, led by the PIs:

PI, Gerry Stahl: PI on the VMT Project. Author of *Group cognition: Computer support for building collaborative knowledge* and *Studying virtual math teams*. Founding editor of *International Journal of Computer-Supported Collaborative Learning*. He will have overall responsibility for the project.

PI, Arthur Powell: Chair of the Department of Urban Education at Rutgers-Newark and Associate Director of the Robert B. Davis Institute for Learning at Rutgers-New Brunswick. Specializes in problem solving, deductive reasoning and heuristics in math education. Expertise in analysis of learning in digital video. Primary responsibility for teacher professional development at Rutgers.

Co-PI, Jason Silverman: Faculty member at the School of Education, Drexel University. Developed and teaches the online masters degree program in Mathematics Learning and Teaching at Drexel. Primary responsibility for teacher professional development at Drexel.


Co-PI, Sean Goggins: Brings a decade of collaborative and social software design and development team leadership. He will be primarily responsible for automated and statistical data analysis.

Co-PI, Michael Khoo: Trained in anthropology, he evaluated components of NSF NSDL digital libraries. He will coordinate the internal formative evaluation component of this project.

Annie Fetter: Co-founder of the Math Forum. Directs the Problem-of-the-Week. Has done professional development and written curriculum for the Geometer's Sketchpad software since it was created. She will be involved in training and mentoring the teachers and coordinating the classroom usage.

Sukey Blanc: Trained in urban anthropology, she studies mathematics and science education, educational equity and school reform. She is Senior Research Associate with Research in Action, a Philadelphia-based non-profit organization engaged in education research and evaluation, which since 1992 has worked with public school districts, educational institutions and community organizations to improve educational
opportunities for those traditionally disadvantaged. She will work with the Advisory Committee and will be responsible for external formative and summative evaluation.

**The Math Forum.** This well established math education site, MathForum.org, has its office at Drexel University with program and technical staff to run services and to maintain the Internet technology. The staff has extensive experience in mentoring math teachers, training new mentors, designing math resources and supporting a huge user community. Most of the program staff are experienced classroom math teachers. The technical staff will be responsible for software development during the project and then for maintaining the project software during and beyond the lifetime of the project.

**The Advisory Committee.** The AC brings expertise in math education; educational psychology; quantitative analysis of learning outcomes, motivation and attitudes; problem-based learning theory and analysis; CSCL; and online communities of learners. (See attached letters.)

**References**


Curriculum and Instruction, Florida International University. Miami, FL. Web: http://digitalcommons.fiu.edu/etd/136


Proposals for Research

in the service of learning. Cambridge, UK: Cambridge University Press.


ONR: Theories and Models of Group Cognition

Statement of Work

This project brings a broad range of theoretical approaches, mixed-method analyses and computational models to bear on a rich data set of team interaction. The data provides a complete record of eight hours of intense synchronous problem solving by two virtual math teams. The data was collected in 2006 under IRB-approved protocols; the data is completely anonymous online chat data.

Within the project, the data will be analyzed in three primary ways: (i) through manual qualitative conversation analysis, (ii) through leading-edge techniques of natural language automated processing and (iii) through mixed methods of qualitative and quantitative analysis, data mining, cluster analysis, statistical analysis and network analysis.

The findings from the original analyses described above will be compared with a range of relevant previous literature. This includes the following sources: (i) previous work of the PIs themselves, (ii) related work by other researchers in the ONR CKI program, (iii) coding schemes developed in the field of computer-supported collaborative learning (CSCL), and (iv) seminal works on distributed cognition, situated cognition, activity theory, mediated cognition, situated learning, knowledge building, ethnomethodology, actor network theory, dialogics, small-group theory and social theory.

In addition to publishing project findings in white papers, conference papers and journal articles, the PIs will organize workshops to: (i) compare different coding schemes and analysis methods and to discuss potentials for synthesis and mixed methods combinations, (ii) analyze the data set for this project from different methodological perspectives from other CKI and CSCL projects, and (iii) consider different theories and models of macrocognition as applied to the data set for this project.

The goals of the project will be: (i) to identify the nature of group cognition processes (macrocognition) in ad hoc problem-solving teams, (ii) to clarify
terminology, (iii) to distinguish related theories, (iv) to validate or expand theory, and (v) to contribute to computational models and other tools, coding schemes and metrics for analyzing macrocognition.

Introduction

This project brings a broad range of theoretical approaches, mixed-method analyses and computational models to bear on a rich data set of team interaction. The data provides a complete record of eight hours of intense synchronous problem solving by two virtual math teams. The data will be made available in a number of formats convenient for analysis. Within the project, the data will be analyzed in three primary ways:

• Using an adaptation of conversation analysis applied to text chat, the interactions will be analyzed to identify methods of group cognition or macrocognition, whereby the group constructs new knowledge that emerges through the group interaction and that none of the participants previously possessed.

• Using techniques of natural language processing, the interactions will be automatically coded using coding schemes that identify key moves and utterances that are associated with collaborative knowledge building or macrocognition.

• Using mixed methods of qualitative and quantitative analysis, data mining, cluster analysis, statistical analysis and network analysis, the two approaches above will be bridged, identifying measures that connect the qualitative manual conversation analysis results and the quantitative automated coding analysis results.

The findings from the original analyses described above will be compared with a range of relevant previous literature. This includes the following sources:

• The previous work of the PIs themselves, including the analyses in Stahl’s *Group Cognition* and *Studying Virtual Math Teams*, and the past work by Rosè on language analysis and coding of knowledge building.

• Related work in the CKI program, including publications from projects funded by CKI, such as Cooke’s and Warner’s analyses of interaction data.

• Other coding schemes for collaborative knowledge building developed in the field of computer-supported collaborative learning (CSCL).
• Seminal works on distributed cognition, situated cognition, activity theory, mediated cognition, situated learning, knowledge building, ethnomethodology, actor network theory, dialogics, small-group theory and social theory.

The project leads (Stahl and Rosé) are both leaders in the international CSCL research community. The project will leverage their connections in the CSCL and CKI communities to involve other researchers in collaboratively pursuing the project investigations and in disseminating the emerging results. In addition to publishing project findings in white papers, conference papers and journal articles, the PIs will organize the following kinds of events:

• A workshop at which researchers with different methodological perspectives from other CKI and CSCL projects gather to compare different coding schemes and analysis methods and to discuss potentials for synthesis and mixed methods combinations.

• A workshop at which researchers with different methodological perspectives from other CKI and CSCL projects gather to analyze the data set for this project.

• A workshop at which researchers representing different theoretical perspectives gather to consider different theories and models of macrocognition as applied to the data set for this project.

The goal of the project will be to identify the nature of group cognition processes in ad hoc problem-solving teams, to clarify terminology, to distinguish related theories, to validate or expand theory and to contribute to computational models and other tools, coding schemes and metrics for analyzing macrocognition.

Impact of Proposed Work

The proposed project will result in the design, development and testing of analysis methods, automated tools, dynamic models and empirically grounded theory for the understanding of group processes of macrocognition (aka group cognition) in ad hoc teams confronted by non-standard problems.
Future Naval Relevance

The proposed project is directly responsive to the ONR CKI Program focus on analyzing group processes involved in team decision making in tactical teams. The project develops tools for analyzing, theorizing and modeling group processes involved in team decision making in small ad hoc groups collaborating on complex problem exploration, analysis and solving.

Management Approach

Gerry Stahl will coordinate work at Drexel and Carolyn Rosé will coordinate work at CMU. They will stay in weekly contact to coordinate the overall project. Drexel will act as lead on the grant and 50% of the grant is subcontracted by Drexel to CMU.

All human data to be used is strictly anonymous online chat data recovered from Math Forum server logs. The data was created in Spring 2006 under protocol approved by the Drexel IRB, which is certified under Human Subject Assurance Number FWA0001852. CMU’s IRB is also certified.

Technical Approach

Outline of Proposed Work

In each of the project’s three years, there will be six types of tasks, including (1) corpus definition, (2) manual analysis, (3) coding scheme design, (4) automated coding, (5) data analysis and (6) theory building:

1. **Corpus Definition:** In each year of the proposed work, we will work with a different existing corpus of interaction data so that by the end of the project, we will be able to engage in theory building that generalizes across multiple tasks under multiple configurations. By the end of the project, we will be in a good position to derive generalizations that have substance and generalizability. We will apply both the transactivity-based coding scheme and coding schemes from the CKI and CSCL communities to the same data.

2. **Hand Analysis:** For each of our corpora we will analyze up to half of the data by hand and then use automatic coding technology to code the rest. This hand analysis will be based upon interaction analysis of the corpus.
Although the hand analysis will attempt to uncover structures to guide the design of the coding scheme, it will more generally seek to discover the full range of macrocognitive processes that take place in the data at the group unit of analysis.

3. **Coding Scheme Design**: Our work will be focused on a transactivity-style analysis, however we expect to have to make adjustments to the category definitions for each corpus we work with in order to be true to the nuances of the discussions going on there while maintaining high reliability and without changing the spirit of the codes. Additionally, we will be working with coding schemes from the CSCL and CKI communities, beginning with Cooke and Gorman’s (2009) work on interaction-based measures of cognitive systems, especially measures of communication flow, which allow for analyses of influence and stability within group discussions.

4. **Automated Coding**: As in our prior work, we will make heavy use of automated coding technology in this proposed work. In our experience, the technology is still new enough that each corpus we work with raises new challenges for this technology. However, as we address those challenges, we produce new knowledge in the area of text mining and text classification, which generates additional insights and publications.

5. **Data Analysis**: One major goal of our data analysis across all three corpora is to validate the transactivity framework by correlating occurrences of subsets of codes with important outcome measures. But we’ll also be exploring correlations between occurrences of transactivity-related events with those of the types of analysis schemes explored previously in the CKI and CSCL communities. In general, the data analysis will explore diverse methods and mixed-method combinations to specify data points and group interaction methods (macrocognitive processes) as discovered in the data by both hand analysis and automated coding, in order to test and refine theories and models of team decision making in ad hoc groups.

6. **Theory Building**: The ultimate goal of our theory building will be to stimulate exchange of ideas and findings between the CKI community and the CSCL community through workshops, symposia and publications at the International Conference of the Learning Sciences, the Computer-Supported Collaborative Learning conference and the *International Journal of Computer-Supported Collaborative Learning*. 
Detailed Description of Project

1. Corpus Definition

Data that captures interesting examples of collaborative knowledge building is hard to find. The Group Cognition Lab worked for six years to generate good data for analysis (Stahl, 2006). It defined an online environment in which groups can meet and everything that group participants share interactionally is captured by the computer logs. We defined tasks and facilitated sessions to realize ad hoc, complex, one-of-a-kind team problem-solving scenarios. We led the groups to focus on building and processing new knowledge for their problem solving. Moreover, we recruited students at a stage where they were just learning the fundamentals of abstract thinking, so that we could observe the emergence of new individual and group skills in concert with each other. The lab developed technologies for instrumenting the online environment and for replaying the interactions in ways that support detailed analysis by researchers. In addition, we explored alternative analysis methods and developed our own approach to interaction analysis.

The core data set selected for this project was generated as part of the Math Forum’s VMT Spring Fest 2006 in May of 2006. The student participants were normal users of the Math Forum online services; their identities were completely anonymous, signified only by a self-selected login chat handle. The best examples of group cognition can be found in the logs of Team B and Team C. These logs reveal rich examples of cognitive processes accomplished interactively by the groups. Each Team engaged in four hour-long sessions during a two-week period. There are dramatic signs of longitudinal development at both the individual and group level as they learned new communication and problem-solving skills and methods appropriate to their socio-technical and goal-oriented situation.

The log for the two teams together consists of about 3,000 chat postings and 3,000 other actions. This is a sizable corpus for manual and automated analysis. We already have considerable experience analyzing brief excerpts from this corpus. These excerpts form the core of two exceptional PhD dissertations that have already been completed (Çakir, 2009a; Sarmiento-Klapper, 2009a). Other excerpts have been analyzed by colleagues from other labs internationally, as reflected in chapters of Studying Virtual Math Teams (Stahl, 2009b) and in symposia on VMT data at the CSCL 2007 and 2009 conferences (Koschmann & Stahl, 2009; Stahl, 2007).
The core data set is being made available as open source through an international CSCL data archive. This will not only make it globally available to researchers for making comparisons, but it will format it in a common XML-based scheme, making it susceptible to being displayed in various templates. This is part of an on-going effort within the CSCL community to enhance comparability of different methodological approaches. The proposed project will be part of this international effort in a number of ways.

The selected data corpus will be analyzed in detail within the proposed project through three phases:

- **Year I: Session 4 of Team B.** This is probably the session with the most examples of collaborative knowledge building. Therefore it will provide a rich source for initial development of a coding scheme that identifies and classifies instances of effective macrocognition.

- **Year II: Team C Sessions 1 and 4.** This is data involving the same web-based technology and the same problem-solving task as in Year I, but conducted by a different group of participants. The inclusion of the team’s first and last session offers data with a longitudinal contrast, as well as some comparison with the year I data. It therefore provides a solid basis for testing and generalizing the year I coding.

- **Year III: The complete combined corpus of Team B and C data (all sessions).** This provides an extensive data corpus of over 6,000 events. It includes many group interactions. It provides a rich source for statistical comparisons among interactions.

### 2, Hand Analysis

The VMT Project at the Group Cognition Lab at Drexel University has developed an ethnomethodologically-informed approach to interaction analysis of synchronous online interaction data (Zemel, Xhafa & Çakir, 2009). This approach is defined and described in Chapter 28 of *Studying Virtual Math Teams* (Stahl, 2009c). It is illustrated especially in Chapters 6, 7, 8 and 9 of that volume (Çakir, 2009b; Sarmiento-Klapper, 2009b; Toledo, 2009; Zhou, 2009). The method involves data sessions using the VMT Replayer to engage a group of experienced researchers in the conversation analysis of an excerpt from an online session to define the linguistic, visual and indexical work being carried out by the group and the group cognition thereby accomplished. The method is rigorous, generalizable and reliable, as discussed in Chapter 28.
As described in Chapter 28 on “Toward a Science of Group Cognition” (Stahl, 2009c), the analysis of group cognition explores how small groups engage as a group (i.e., at the group unit of analysis) in the accomplishment of cognitive tasks. These include such tasks as: intersubjective meaning making, interpersonal trains of thought, shared understandings of diagrams, joint problem conceptualizations, common references, coordination of problem-solving efforts, planning, deducing, designing, describing, problem solving, explaining, defining, generalizing, representing, remembering and reflecting. Groups develop general methods of doing these things, always adapted to the situations in which they are engaged and the media and other resources that are at their disposal (Stahl, 2009a).

3. Coding Scheme Design

Machine-learning algorithms can learn mappings between a set of input features and a set of output categories. They do this by using statistical techniques to find characteristics of hand-coded “training examples” that exemplify each of the output categories. The goal of the algorithm is to learn rules by generalizing from these examples in such a way that the rules can be applied effectively to new examples. In order for this to work well, the set of input features provided must be sufficiently expressive, and the training examples must be representative. Typically, machine-learning researchers design a set of input features that they suspect will be expressive enough (Strijbos, 2009). At the most superficial level, these input features are simply the words in a document. But many other features are routinely used in a wide range of text-processing applications, such as word collocations and simple patterns involving part of speech tags and low-level lexical features; we will draw from this prior work.

Once candidate input features have been identified, analysts typically hand code a large number of training examples. The previously developed TagHelper tool set (Rosé et al., 2008) has the capability of allowing users to define how texts will be represented and processed by making selections on the GUI interface. In addition to basic text-processing tools such as part-of-speech taggers and stemmers that are used to construct a representation of the text that machine-learning algorithms can work with, a variety of algorithms from toolkits such as Weka (Witten & Frank, 2005) are included in order to provide many alternative machine-learning algorithms to map between the input features and the output categories. Based on their understanding of the classification problem, machine-learning practitioners typically pick an algorithm that they expect to perform well. Often this is an
iterative process of applying an algorithm, seeing where the trained classifier makes mistakes, and then adding additional input features, removing extraneous input features, or experimenting with algorithms.

Applying this iterative process requires insight and skill in the areas of linguistics and machine learning that the social scientists conducting corpus analysis are unlikely to possess. TagHelper tools support this interactive processes by making it easy to define different processing configurations through the GUI and then providing reports about how the configuration worked and where the process may have broken down. The goal of our tool development is to make this process easier for social scientists. In particular, the process of identifying where the process has broken down and how the configuration can be tuned in order to improve the performance requires more expertise than typical social scientists would possess. Thus, the bulk of our development work will be in developing the machinery to bridge the gap between the natural structure of the input texts and the behaviors that social scientists are interested in cataloguing and coding, using bootstrapping approaches.

In our recent corpus-based experiments (Josh & Rosé, 2009; Arora, Joshi, & Rosé, 2009) we have explored the usage of alternative types of syntactically motivated features on text classification performance. Our methodology is extensively discussed in our recent journal article in the *International Journal of Computer-Supported Collaborative Learning*, investigating the use of text classification technology for automatic collaborative learning process analysis (Rosé et al., 2008).

**Advancing Beyond the Capabilities of Keyword-Based Approaches.**
Linguistic Inquiry and Word Count (Pennebaker, 2003) is a paradigm case of keyword-based approaches to analysis of verbal data, that is very commonly used in social psychology, especially but not solely in work related to health communication. LIWC indicators that are designed to measure latent characteristics of authors such as emotional or psychological state based on vocabulary usage have been successfully calibrated with a wide range of behaviors over multiple types of studies. Nevertheless, they have limitations that must be taken into account methodologically. LIWC indicators have typically been used in studies where the external variables of interest are health outcomes or health related behavior. In studies where consistent stories based on calibrations of LIWC indicators with external variables are reported, the corpora used were created under very controlled circumstances, always only within the experimental condition of a study in which the genre and topic of the writing were determined by the experimental manipulation. When these tight constraints are removed, the
story becomes much less clear. For example, Pennebaker and Francis (1996) present results from a study with two different conditions. The experimental variation lay in the change of the topic participants wrote about. In this study, the LIWC indicators made opposite predictions about behavioral outcomes and emotional states in the experimental condition in comparison to the control condition. Discrepancies like this occur because there are many linguistic factors besides the emotional state of the author or speaker that affect the frequencies of word usage. For example, many words have multiple meanings and only convey negative emotion in some contexts and not in others. For example, the words “bake” and “roast” used while talking about the weather convey a feeling of discomfort, whereas in the context of a discussion about cooking, they do not. Base frequencies of terms also vary between topics. Thus, a difference in frequency of a term may either indicate a difference in the emotional state of the author or simply a difference in topic. If LIWC predictors were truly indicative of emotional state independent of topic, and fluctuations in emotional state predict corresponding fluctuations in health and behavior outcomes, it is difficult to reconcile the difference in the direction of predictions between conditions reported in that paper. Nevertheless, if one accepts that LIWC indicators are merely proxies that can be used for estimating measurement of psychological state within very narrowly constrained contexts, then the pattern makes sense. However, this limitation has strong negative implications for the applicability of LIWC indicators within naturalistic communication settings in which there is a wide variation in the communicative goals motivating individual contributions, such as in naturalistic on-line learning environments where students may interact about a wide range of topics in connection with a variety of activities over time.

Analysis of collaborative learning interactions have demonstrated that what happens on the process level is important for predicting what cognitive benefits participants in a conversation take away from it (e.g., King 2007). More complex learning is supposed to occur in “spirals of reciprocity,” where learners are intensely engaged with one another (Salomon and Perkins 1998). In particular, learners can attain new levels of understanding during interactions where more complex cognitive activities occur, such as analytical thinking, integration of ideas and reasoning. These include activities such as elaborating on content (e.g., Webb 1989), explaining ideas and concepts (e.g., Chi et al. 1994), asking thought-provoking questions (e.g., King 1998, 1999), argumentation (e.g., Kuhn 1991), resolving conceptual discrepancies (e.g., Piaget 1985) and modeling one another’s cognitive states. These activities may not be adequately represented by patterns of individual turns taken out of context. Modeling these processes
instead requires categorical coding schemes building on precise definitions of categories (see Chi et al. 1994). Trained human coders are able to consistently apply well-defined coding schemes across multiple contexts. However, we acknowledge that applying coding schemes like this by hand is extremely tedious. And effectively writing rules by hand to reliably match against complex patterns, which is an option provided by some corpus analysis environments, is difficult as well.

When human coders apply categorical coding schemes, they bring insights with them from their human intellect. Human language is highly complex, encoding meaning on multiple levels, and carrying very subtle nuances that are difficult to formally capture with a rule based model. Interpretation of language involves using cultural sensitivity to style and lexical choice, applying world knowledge, integrating meaning across spans of text, and often making inferences about what is implied in addition to what is literally stated. In contrast, regardless of approach, machine coding will always be based on rigid rules that are necessarily an over-simplification of the reasoning processes that humans rely on for their interpretation. Note that word counting approaches such as LIWC, which were discussed earlier, are an extreme case of this over-simplification. This simplification threatens the face validity of the coding that can be accomplished automatically because this word based approach may not be measuring what it is purported to be measuring. Using an example from our own work, we have used LIWC to examine the language behavior of five different tutors who participated in a series of calculus problem solving studies (Gweon et al. 2006). We evaluated tutor effectiveness by comparing them with respect to the average learning gains of the students they tutored. Based on this analysis, we determined that the more effective tutors scored higher on LIWC’s confidence scale. When we examined which words from the tutors’ contributions the associated LIWC word list was matching against, the most frequent word was “factor”, which came up inside discussions about algebra. Thus, the LIWC confidence scale was not ranking tutors based on their confidence at all, but rather their tendency to supplement their calculus tutoring with basic algebra concepts such as factoring. Thus, word-counting approaches like LIWC that make their assessment based on individual words taken out of context should be used with caution. We see from our calculus example that they are not guaranteed to reflect accurately the mental states they were designed to assess.

Machine learning based approaches can transcend the limitations of keyword-based approaches because they allow for more complex representations of text beyond simply keywords. In our recent work, for example, using more complex grammar and context oriented features in
addition to word level features, we have demonstrated significant improvements in analysis accuracy over simple word based representations of text for tasks such as collaborative learning process analysis (Rosé et al., 2008), sentiment analysis (Joshi & Rosé, 2009; Arora, Joshi, & Rosé, 2009), and text compression (Chaudhuri, Gupta, Smith, & Rosé, 2009; Gupta, Chaudhuri, & Rosé, 2009).

**Advancing Beyond the Capabilities of LSA.** Latent semantic analysis (LSA) is well known as a practical method for representing words in terms of classes of words that share a similar distribution in terms of the “neighborhoods” or words they occur with. One can think of it as a way of identifying groups of semantically related words (Landauer et al., 1998). We would expect that methods that offer a way of generalizing over alternative phrasings of the same or similar ideas would be useful in tracking initiation-reply links that form the elementary units of knowledge building processes in conversation.

In the typical method for applying LSA, we first construct a term-by-document matrix. Next, LSA applies singular value decomposition to the matrix, and reduces the dimension of the feature space of terms to a 300-dimensional concept space. We can then represent a term vector, whether it is a simple term vector or an expanded term vector, in terms of this LSA space by averaging across the LSA representation for each word in the text within that 300 dimensional space. Text vectors that have been transformed in this way can then be compared using cosine similarity.

However, for the purpose of tracing the knowledge-building process of students, there is a major limitation of LSA as it is typically applied that must be taken into account. Note that not all words carry equal weight within the vector that results from the averaging process in constructing an LSA vector for a text. Words that are closer to the “semantic prototypes” represented by each of the 300 dimensions of the reduced vector space will have vectors with longer lengths than words that are less close to any single one of those prototypes within that space. Thus, those words that are closer to those prototypes will have more of an effect on the direction that the resulting vector will have within the space. Thus, they will have more of an effect on the comparison with other texts. However, one should note that in a running discussion, it is the unusual content, the noteworthy ideas, that often form the links between initiation and responses, rather than the common concepts that form the background for the ongoing discussion. And thus, one major limitation of LSA as it is typically applied is that it de-emphasizes the contribution of precisely those words that are most
important for making the textual links in the discussions that we would like to identify.

Recently we have developed a new approach to applying LSA that overcomes this limitation. For the task of identifying initiation-reply links in a conversational thread recovery task, it significantly outperformed the typical method for applying LSA as well as other baseline approaches making use of lexical resources such as Wordnet. Further work along these lines will be a major focus of the technical component of this proposed research.

### 4. Automated Coding

Many of the fundamental activities in on-line organizations, such as brainstorming, decision-making and training, require communication. This underlying conversation both furthers the goals of a team and reflects the underlying structure of interactions and relationships within social institutions (Zimmerman & Boden, 1991). Several decades of research in Computer Mediated Communication (CMC) have examined how the use of media affect team communication processes (e.g., Hall, 1976; Li, 1999; Setlock & Fussell, 2007). However, progress in this research community is hampered by how time consuming it is to do this analysis by hand. For example, one recently published study of the effects of culture on negotiation processes required over a year to collect the data, refine the coding scheme, and code and analyze the data. The outcome of this work is a better understanding of one of many communication processes in virtual teams performing one of many different tasks. To extend such work to the full domain of teams, tasks and communicative activities would take decades. As a basic part of our approach, we propose to use a traditional approach to using analysis of corpora by hand in order to increase understanding of virtual teams on a small to medium scale and then use the automatic analysis to expand to a dramatically larger scale.

In our prior work, we have made substantial progress towards detecting properties of conversation that are specifically associated with quality of collaboration. We have focused on a property known as transactivity (Rosé et al., 2008; Wang, Rosé, & Joshi, 2007; Joshi & Rosé, 2007), an important property of collaborative discourse. Participants in a collaborative setting are said to have transactive discussions when they elaborate, build upon, question or argue against the ideas presented by their partners in the process of working towards a common understanding of the task and reaching a shared solution. This process of understanding the partners’ ideas,
comparing them to one’s own understanding, arguing and forming a common ground upon which a solution can be built collaboratively has been shown as important for learning (Teasley, 1997; Azmitia & Montgomery, 1993).

The idea of transactivity has its roots in educational psychology with Piaget’s model of assimilation/accommodation and Vygotsky’s socio-cultural theory of learning. Piaget’s model is a particularly key component of the theoretical underpinnings of our proposed work since it provides a framework for characterizing the difference between simply managing existing pieces of knowledge, as one might characterize work so far on macrocognition in the CKI community, and more major cognitive restructuring that can occur at certain points within an assimilation/accommodation cycle.

Digging into the details a little more deeply, at the heart of Piaget’s theory of learning is the assimilation-accommodation cycle in which students encounter stimuli in the world that provide data either in support of or in conflict with their own internal model of the world. During assimilation, a student imposes his own model on the stimuli he sees, interpreting everything in that light, and rejecting what does not fit. During accommodation, a student is open to considering a model of the world that may be imposed from the outside. When these processes are in balance, the conditions are most favorable for a student to notice a gap or deeper flaw in his own mental model. When a student becomes aware that stimuli from the world reveal a gap, that student may then choose to search for a revised model of the world that accounts for the new data, which may even require a dramatic paradigm shift or major reorganization of knowledge.

It is important to note that an important ingredient in Piaget’s theory is the equality of power or partnership between students working together that is important for creating an environment in which assimilation and accommodation are in balance. Note that equal power does not imply equality in knowledge. Based on Piaget’s foundational work (Piaget 1985), one can argue that a major cognitive benefit of collaborative learning is that when students bring differing perspectives to a problem-solving situation, the interaction causes the participants to consider questions that might not have occurred to them otherwise. This stimulus could cause them to identify gaps in their understanding, which they would then be in a position to address. This type of cognitive conflict has the potential to lead to productive shifts in student understanding. Examining the discourse between students in a collaborative-learning setting can reveal evidence of the power relationship between students, the exchange of views and evidence of the
opportunity for cognitive conflict in the socio-cognitive conflict that is manifested in the argumentation that occurs between students. The impact of socio-cognitive conflict on learning has been noted especially in connection with difficult-to-learn content (Azimitia & Montgomery, 2005; Russell, 2005). And the important connection between relationship development and socio-cognitive conflict has also been documented (Azimitia & Montgomery, 1993). Examining the discourse between students can also reveal where an imbalance in a power relationship can hinder participation and learning. For example, Elbers & de Hann (2004) provide a qualitative analysis from a socio-cultural perspective on how racial stereotypes affect the power/authoring relationship between students, which may hinder collaborative discussion.

Vygotsky’s theory argues for similar patterns of discussion from another angle. While Piaget’s theory focuses on equal power but difference in knowledge, Vygotsky focuses more directly on differences in knowledge, but also argues in favor of relationship development and the social nature of knowledge construction. Based on Vygotsky’s seminal work (Vygotsky 1978), we know that when students who have different strengths and weaknesses work together, they can provide support for each other that allows them to solve problems that would be just beyond their reach if they were working alone. This makes it possible for them to participate in a wider range of hands-on learning experiences. In our own work, we have observed evidence of helping behavior as a socio-cognitive variable that mediates learning (Gweon et al., 2006; Gweon et al., 2007). Social aspects of group functioning as they relate to and result from patterns of interaction are unquestionably key consideration for groups that will interact with one another over a long period of time. However, one could argue that they are even more essential in ad hoc teams with a critical purpose since any subtle incident that might harm trust or hinder the flow of information might interfere with the success of the encounter.

Surveying the field of computer-supported collaborative learning for frameworks for analyzing group conversations, one might conclude that there are a plethora of different approaches. Nevertheless, one might also consider it not a giant leap to consider that the topic of what makes group discussions productive for learning and community building has been explored with very similar findings, perhaps with subtle distinctions, and under different names such as transactivity (Berkowitz & Gibbs, 1983; Teasley, 1997; Weinberger & Fischer, 2006) in the cognitive learning community and uptake (Suthers, 2006), group cognition (Stahl, 2006) or productive agency (Schwartz, 1998) in the socio-cultural learning community. Despite differences in orientation between the cognitive and
socio-cultural learning communities, the conversational behaviors that have been identified as valuable are very similar. Building on these common findings, the field of Computer-Supported Collaborative Learning has emerged where support for collaborative learning has been developed that addresses observed weaknesses in conversational behavior related to this phenomenon.

5. Data Analysis

To complement the ethnomethodologically informed interaction analysis and the machine-learning algorithms for automated coding (Strijbos, 2009), we will also analyze the three corpora using content analysis (Krippendorf, 2004) and network analysis (Wasserman and Faust, 1994). The content analysis will be performed on the 3,000 chat postings and the network analysis will be performed on the 6,000 chat and drawing actions.

The content analysis will be executed using two rubrics (Goggins & Laffey, forthcoming). The unit of analysis for this work will be a complete unit of conversation (Krippendorf, 2004). The first rubric will evaluate the development of group identity within the small groups, using Tajfel’s (1978, 1979, 1982) description of group communication as inter-group, inter-personal, intra-group and inter-individual. Inter-group communication is communication across groups, and only rarely occurs in this data set. Inter-personal communication takes place between two individuals. Intra-group communication is within the group, where all members participate in the dialogue. An utterance addressing individual members in the presence of the whole group as an aside is coded as inter-individual communication.

The second rubric will evaluate the corpus of data for knowledge co-construction using a rubric developed by Gunawardena et al (1997). Two raters will score the conversations on these rubrics and measure inter-rater reliability using Krippendorf’s alpha (2004). This type of analysis is performed by Goggins, Laffey & Gaylen (forthcoming) on asynchronous communication records, and the contrast with the results from synchronous chat data will provide a helpful contrast of synchronous and asynchronous knowledge co-construction in small groups.

Social network analysis will be performed on the 3,000 chat postings and 3,000 other actions in order to determine if there are patterns of networked interaction that correspond with the development of group identity or the co-construction of knowledge. The resulting networks will be bi-partite (users and objects) and regular. Since the networks in our corpora are closed and
small, we will focus our analysis on small network evolution and elaborating semantically meaningful measures of tie strength.

Tracking longitudinal evolution will involve developing a time-series set of network views, possibly addressing the state of the network as a feature that contributes to other forms of analysis. We will also explore the advantages of deriving measures of tie strength from the results of machine-learning algorithms, response-time lag and length of sustained interaction between pairs of group members.

These quantitative analyses will not be performed in isolation from the interaction analysis or the automated coding. Decisions about the granularity in both network analysis and content analysis will take the findings and approaches from these other two methods into consideration. The findings of all these mixed-method analyses will inform the design of computational models (Wee & Looi, 2009) and supply a basis for calibrating the models of macrocognition.

6. Theory Building

The findings of the analyses described above will be synthesized into a theoretical framework of group cognition / macrocognition. This theory will be compared to competing theories in current research literature, such as: distributed cognition (Hutchins), situated cognition (Suchman), activity theory (Engeström), mediated cognition (Vygotsky), situated learning (Lave), knowledge building (Scardamalia & Bereiter), ethnomethodology (Garfinkel), actor network theory (Latour), dialogics (Wegerif), small-group theory (Weick, 2005) and social theory (Giddens, 1984). The comparison will aim to determine areas of overlap, respective limitations, potential conflicts and possibilities for synthesis.

Project Schedule, Milestones and Reports

As detailed above, in each of the three years, there will be six types of tasks, including (1) corpus definition, (2) coding scheme design, (3) hand analysis, (4) automated coding, (5) data analysis, and (6) theory building. These six types of activities are broken down into tasks associated with target dates within the three years of the proposed work in the table below.

Deliverables for this project include 5 coding manuals, 3 coded corpora (all of which are coded with two different frameworks, one transactivity based
and one based on Nancy Cooke’s work), and publications (at least 2 submissions per year, which include both conference papers and a journal article and include technological innovation as well as theory building). Extensions to automatic coding technology will be integrated with the already publically available text mining toolkits, TagHelper tools and SIDE, which have been developed in PI Rosé’s prior work and are already in broad distributions (for example, TagHelper has over 1000 users in 57 countries).

**FY2010**

- Chapter on linguistic analysis of collaboration for the *International Handbook of Collaborative Learning* (already in progress)
- Workshop at Alpine Rendezvous on coding schemes for collaborative knowledge building
- Workshop at International Conference on the Learning Sciences on coding schemes for collaborative knowledge building
- Coding manual for Corpus 1
- Coded corpus 1
- Conference paper with automatic analysis results on coded corpus 1
- Quarterly Technical and Financial Progress Reports

**FY2011**

- Coding manuals for corpus 2
- Coded Corpus 2
- Workshop at GROUP on mixed methods for analyzing collaborative knowledge building
- Workshop at CSCL on theories related to macrocognition
- White paper on theories related to macrocognition
- Possible conference paper (ACL or AIED) related to automatic analysis
- CSCL 2011 paper introducing the CKI framework to the CSCL community, with theory building analysis from year 1 on corpus 1
- Quarterly and Annual Technical and Financial Progress Reports
FY2012

- Coding manuals for corpus 3
- Coded Corpus 3
- Publication of book on interaction analysis of Corpus 3 by MIT Press
- ICLS paper submission on results from study 1/Corpus 2
- CHI paper submission for study 2/Corpus 3
- Journal article submission synthesizing findings across all corpus analysis projects for this grant
- Quarterly Technical and Financial Progress Reports
- Final Report

Qualifications of the Principal Investigators

The Group Cognition Lab at Drexel

The Group Cognition Lab conducts basic research on phenomena of distributed cognition that take place distinctively at the small-group level of description, such as collaborative knowledge building, joint decision making, group problem solving, shared meaning making, co-construction of knowledge representations. It is located at Drexel University in Philadelphia and is a joint project of the iSchool (College of Information Science and Technology) and the Math Forum. It is directed by Gerry Stahl, Sean Goggins and Stephen Weimar.

The Lab specializes in studies that make visible the development of group cognitive processes by generating, capturing and analyzing naturalistic episodes of computer-mediated interaction by novices, such as teams of students just learning to problem solve together online. The microanalysis of these episodes reveals characteristics of group process that contribute to an empirically grounded theory of group cognition, which is emerging from the lab.

The Lab is a flexible collaboration of researchers who bring complementary skills and interests to the multidisciplinary mission of the Lab. This includes information scientists interested in small-group cognitive processes,
educators interested in how to promote learning of group-cognitive skills, qualitative and quantitative analysts interested in adapting social science research tools to the analysis of group cognition, software designers interested in developing online environments to support effective collaboration, and theorists interested in elaborating the theory of group cognition.

The following major activities are integrated within the Lab:

• Developing the Virtual Math Teams (VMT) service at the Math Forum for generating real-world data on small groups of students learning to engage in online problem solving of open-ended ill-structured and wicked math problems.

• Working with schools of education and math-teacher-training programs to involve teachers and students in exploring the potentials of the VMT service.

• Conducting collaborative data sessions of researchers to analyze the group interactions taking place in logs of online group work.

• Developing case studies and quantitative analyses of the data from logs of online group work to describe characteristics of group cognition.

• Designing new features for the VMT environment to support group-cognitive accomplishments, based on the microanalysis of interesting cases of usage.

• Extending the theory of group cognition, including building graphical and computational models, clarifying terminology, defining specific concepts, and relating to cognate theories.

The Lab has been recognized as a leading center for research on group cognition based on its work from September 2003 to August 2009. It has gone through many cycles of design-based research using a prototype VMT environment at the Math Forum, including Spring Fests in 2005, 2006 and 2007, in which student groups from around the world met for sequences of four hour-long sessions. This produced 2,000 student-hours of data, which was reported in about 200 academic publications. In addition, two major books were published: Group Cognition (Stahl, 2006, MIT Press) assembled studies of online collaboration that motivated the work of the Lab and the VMT service; Studying Virtual Math Teams (Stahl, 2009, Springer Press) includes the most important reports from the Lab and from collaborating researchers.

Potential directions for the coming years include the following:
• Design and implement additional functionality for the VMT collaboration environment, including dynamic geometric representations and intelligent tutoring support. (Research question: How do visual representations and automated guidance contribute to establishing common ground and scaffolding problem solving?)

• Explore web interfaces to support the spontaneous formation of ad hoc virtual teams within a large distributed community, including participants from different cultures and different time zones. (Research question: How to stimulate and support ad hoc teams and how to overcome geographic or cultural differences?)

• Further integrate synchronous and asynchronous media to coordinate group accomplishments at different time scales and different social scales, from intense interaction of small groups to community knowledge building over years. (Research question: What differences do temporal and social scales introduce into group cognition? How to archive synchronous interaction content as useful knowledge and data for the community to reuse asynchronously?)

• Scale up the VMT service to be a regular, year-round service of the Math Forum, used by a large number of groups in creative ways. (Research questions: How to foster and support an online community with minimal staffing, and to manage large numbers of interactions within a safe and productive context?)

• Collaborate with teachers and with math-teacher training programs to enhance the pedagogy, to support teacher involvement and to extend the user base of the VMT service. (Research questions: How to build a distributed community with different levels of expertise and to build teacher's reflective practice through participation in VMT?)

• Continue to hold data sessions of researchers to analyze data from new usage and to explore phenomena of interest in more depth. (Research question: What are the characteristics of group-cognitive problem-solving processes?)

• Apply new qualitative and quantitative social-science methods to the analysis of group-cognitive phenomena. (Research question: How to combine, e.g., conversation analysis and social network analysis or automated coding?)

• Develop quantitative measures of social presence, task performance, cooperative practices, longitudinal social relations and collaborative information behavior in self-assembling synchronous/asynchronous
teams. (Research question: How can we measure processes on online group cognition?)

- Conduct a longitudinal microanalysis of the entire transcript from two four-hour Spring Fest sessions. This would be a ground-breaking analysis approach and an innovative style of monograph, to be published by MIT Press. (Research question: What are the methodological issues in moving from diachronic snapshots of group cognition in brief excerpts to longitudinal changes in collaboration and shared understanding?)

- Continue to publish analyses and to share data with international collaborators. Further refine the theory of group cognition, including building graphical and computational models. (Research question: How can aspects of the theory be summarized in models?)

It is important to note that these aspects of future work are not separable, but need to be conducted as parts of the integrated work of the Lab. The foundational theoretical work of the lab builds upon empirical microanalysis of situated practical activities and aims to contribute to the improved design of tools, concepts and principles to support practical activities.

Gerry Stahl is a leading researcher and theoretician in computer-supported collaborative learning (CSCL). He has presented at every CSCL conference and founded the International Journal of CSCL. Trained in computer science, human-computer interaction, artificial intelligence, cognitive science and philosophy, he is a tenured Associate Professor at the College of Information Science & Technology, Drexel University.

Sean Goggins specializes in mixed-methods research on virtual teams. He is an experienced software developer and is now Assistant Professor at the College of Information Science & Technology, Drexel University.

Stephen Weimar has been the Director of the Math Forum at Drexel University for 15 years. The Math Forum is the premier online resource for mathematics, receiving more than three million visits monthly.

The Group Cognition Lab includes other faculty, graduate students and staff at Drexel and elsewhere, including specialists in anthropology, conversation analysis, educational psychology, math education, teacher training and computer science. The Lab has on-going collaborations at Carnegie Mellon University, Rutgers, Hawaii, Missouri, Wisconsin, Singapore, Germany, Brazil and Romania.
The Language Technologies/HCI Institutes at CMU

Carolyn Rosé holds a joint appointment between the Language Technologies Institute and the Human-Computer Interaction Institute at Carnegie Mellon University (CMU). Locally at CMU, between her two departments she supervises or co-supervises a group of 10 graduate students, a post-doc, and a small number of undergraduates. As a tenure track professor at Carnegie Mellon University, she teaches courses related to collaborative learning, linguistic analysis, machine learning & text mining, and summarization.

The School of Computer Science at Carnegie Mellon University consists of 6 departments, including the Computer Science Department, the Machine Learning Department, the Language Technologies Institute, the Software Engineering Institute, the Human-Computer Interaction Institute, the Robotics Institute. The Language Technologies Institute is the only department of its kind in the nation that is completely dedicated to research in language technologies, and includes researchers from the full gamut of areas within that field. Similarly, the Human-Computer Interaction Institute was the first department of its kind and one of only two universities in the US with such a large and diverse faculty spanning all areas of the field of Human-Computer Interaction and containing the largest number of faculty named as CHI Fellows of any institution in the nation.

Carolyn Rosé is the Co-leader of the Social and Communicative Factors in Learning thrust of the Pittsburgh Science of Learning Center, which includes over 40 faculty from a variety of departments including Psychology, Education, Language Technologies, Robotics, and Human-Computer Interaction, both at Carnegie Mellon University and the University of Pittsburgh who are doing learning sciences research. The confluence of Rosé’s two departments and the Pittsburgh Science of Learning Center provide a unique combination of human and technological resources that make her imminently well situated to successfully carry out innovative research.

Building on a foundation of research in speech translation, dialogue systems, intelligent tutoring, robust language understanding, and machine learning, Rosé has been working in the area of automatic discourse analysis for the past 15 years and has produced 25 peer reviewed publications related specifically to this topic (in addition to over 60 other peer reviewed publications on other topics) in prestigious venues such as the International Journal of Computer-Supported Collaborative Learning, the Proceedings of Computer Supported Cooperative Work, the Proceedings of the Association...
for Computational Linguistics, the Proceedings of Artificial Intelligence in Education. She was recently invited to write a chapter on linguistic analysis of collaborative learning for the International Handbook of Collaborative Learning. As the Secretary/Treasurer of the International Society of the Learning Sciences, Rosé has great visibility in the computer-supported collaborative learning community, and has co-organized workshops related to analyses of collaborative learning discussions yearly for the past four years. She is leading a number of research efforts, including a project bringing together research from the computer-supported collaborative learning community with that of the classroom-discourse community in collaboration with Lauren Resnick at the University of Pittsburgh, a project related to analysis of intercultural communication with Susan Fussell at Cornell University, and a project related to dynamic support for virtual math teams with Gerry Stahl at Drexel University. As a product of an earlier ONR funded effort, Rosé produced the TagHelper tools package for text mining that currently has over 1100 users in 58 countries.

References


Kumar, R., Gweon, G., Joshi, M., Cui, Y., Rosé, C. P. (2007b). Supporting students working together on math with social dialogue. *Proceedings of the SLaTE Workshop on Speech and Language Technology in Education.*


perspectives on peer learning (pp. 179-196), Mahwah, NJ: Lawrence Erlbaum Associates.


ALT: Dynamic Support for Virtual Math Teams

On-line learning promises education for the masses – quality educational opportunities available to all people, but especially those who are in the greatest need – although this dream is yet to be made a reality. The long term goal of the proposed work is to replicate the impact of local, on-campus programs targeting increased college preparedness and college success of minority and low income students, such as the well known Treisman Berkeley Professional Development Program, in a freely available, on-line learning environment. Making what would normally be a staff intensive program available ubiquitously at a dramatically reduced expense would be an enormous payoff. Our proposed solution is to develop a technological augmentation to available human support in a lightly staffed environment. This proposed project brings together a team with expertise in both technological development and careful experimentation both in the lab and in the classroom, a track record for large scale deployment of educational materials, a solid foundation in significant results from prior work on which we build in the areas of computer supported collaborative learning and tutorial dialogue systems.

*Intellectual Merit:* Our research attempts to understand how to structure interactions among peer learners in online education environments to address these problems. The proposed project seeks to enhance participation and learning in the Virtual Math Teams (VMT) online math service by designing, developing, implementing, testing, refining and deploying computer-supported tools to support facilitation in this lightly-staffed service. The key research goal is to optimize a design and implementation for adaptive feedback in support of collaborative problem solving that will maximize the pedagogical effectiveness of the collaboration by eliciting behavior that is productive for student learning in collaborative contexts.

*Broader Impact:* We are working towards understanding the pedagogical and technological features that make on-line education in general, and collaborative learning in particular, effective. If we can understand the causal connections between interaction and learning, then we can wield technology in ways that achieve maximal cognitive and social benefits for
on-line learners. To the extent we are successful, our research will help realize the promise of on-line learning. Expensive instructors and content providers will continue to develop course materials and act as moderators to the extent that resources allow. Fellow students will support each other in dealing with their struggles with the materials. Inexpensive machine agents will aid human facilitators both in matching students who can help each other and as well as by offering help to structure their collaborative learning conversations so that the presence of other students will lead to greater commitment to the course and learning. Making free open courses more successful will benefit all students, but especially those with lower income and from developing countries.

1. Vision

American children are in the middle of a group of 38 countries in terms of science and math education, far behind such countries as Singapore, Korea, Hong Kong or Japan (Mullis et al., 2000). On-line learning offers the potential to address this problem by providing free or inexpensive supplementary education for the masses – quality educational opportunities ubiquitously available, especially those who do not have the resources to pay for high quality private tutoring for their children. While this vision does not address the problem that some of the neediest students do not have access to computer resources, this vision is in line with the Advanced Learning Technologies mission to enable radical improvements in learning through innovative computer and information technologies.

The ultimate goal of the proposed work is to replicate the impact of what are normally local, on-campus programs targeting increased college preparedness and college success of minority and low income students, such as the Professional Development Program (PDP) (Treisman, 1985), in a freely available, on-line learning environment. We focus on middle school math since middle school is a pivotal time when students, especially girls, begin to lose confidence in and interest in math (Callahan & Clements, 1984; Dossey, Mullis, Lindquist, & Chambers, 1988; Brandon & Newton, 1985), and we target the well established Virtual Math Teams (VMT) online math service at http://mathforum.org/vmt as a venue for broad dissemination because of its strategic location in an on-line math service that reaches millions of students per week. In supporting collaboration, we focus on eliciting productive helping behavior, which we have observed to mediate learning in prior studies with this age group and domain content area (Gweon et al., 2007) as well as studies with older students (Gweon et al.,
Furthermore, we focus on eliciting proof-like explanations from students as part of our support for their helping behavior, since this is an important skill connected with a deep understanding of math concepts, and which continues to be a struggle for students throughout their school years. We bring together a team with expertise in technological development, careful experimentation in the lab and in the classroom as well as insightful ethnographic research in real on-line learning environments, a track record for large scale deployment of educational materials, and a solid foundation in significant results from prior work on which we build in the areas of computer supported collaborative learning and tutorial dialogue systems.

The purpose of this project is to enhance participation and learning in the Virtual Math Teams (VMT) online math service by designing, developing, implementing, testing, refining and deploying computer-support tools to enhance facilitation that is available to support students in this lightly-staffed service. It is the lightly staffed nature of this service that makes it a more economical solution that on campus programs such as PDP, mentioned above. One key research goal is to optimize the design and implementation of dynamic collaborative learning support agents that will participate in VMT chat sessions in order to maximize the pedagogical effectiveness of those interactions. Prototype dynamic support agents have already yielded positive learning effects in our pilot evaluations in lab (Wang et al., 2007) and classroom studies (Kumar et al., 2007-a; Chaudhuri et al., to appear) in the domains of science and engineering respectively, and a recent pilot evaluation shows promise with middle school kids learning about fraction arithmetic (Kumar et al., 2007-b). Another key research goal is to develop technology for monitoring collaborative behavior and automatically generating reports for human facilitators to allow them to quickly identify teams that require more attention (Kang et al., to appear-a; Kang et al., to appear-b). Our recent work on automatic collaborative learning process analysis from collaborative learning discussions between college age students (Donmez et al., 2005; Wang et al., 2007c, Rosé et al., in press) provides a foundation for this. In our proposed work we will carry this further by identifying which conversational events are most indicative of a need for support in interactions involving middle-school kids, who are less sophisticated in their communication skills and thus struggle with different issues in collaborative contexts. This will be accomplished through collaboration among CMU, Math Forum and VMT researchers.

We have already begun our joint work by integrating our research findings and infrastructure from our prior work in the areas of computer supported collaborative learning and tutorial dialogue systems. We have also piloted our integrated VMT environment, which we refer to as VMT-Basilica.
Proposals for Research

(Kumar et al., submitted-a; Kumar et al., submitted-b), in a purely on-line setting in order to collect realistic development data and so that our plans for our continued collaboration can be strongly influenced by observations of interactions in the exact environment where we will do our most important work towards a significant impact in the long run. In our exploratory data analysis we have taken a hybrid qualitative/quantitative approach to get a firm handle on consistent patterns that are general across the data as well as to notice the influence of important contextual variables that we will take into account in our subsequent experimental work, in line with methodology proposed in (Design-Based Research Collective, 2003).

Our research goal is supporting productive collaborative learning discussions in a computer-mediated environment in “the wild”, specifically supporting students in working together in pedagogically effective ways. While the help students are capable of offering one another is not perfect, there is evidence that it is effective in spite of the errors students make when helping each other (Gweon et al., 2006), and possibly even because of these errors (Piaget, 1985; De Lisi & Goldbeck, 1999; Grosse and Renkl, submitted). If we can harness the potential of state-of-the-art technology for automatically filtering collaborative learning discussions that we have developed in our previous work (Donmez et al., 2005; Wang et al., 2007c), and we can use this automatic analysis to trigger interventions that support students in helping each other learn together (Gweon et al., 2006) using tutorial dialogue and intelligent tutoring technology as in some of our previous studies (Wang et al., 2007; Kumar et al., 2007), we could move towards a solution to our nation’s educational problems in a cost effective, practical manner. To this end, our main research objectives include:

1. Extending the capabilities of the technical infrastructures created in our prior work at Carnegie Mellon University and Drexel University, which includes an elaborate environment for coordinating math teams and supporting their problem solving efforts as well as tools for automatic collaborative learning process analysis and for building collaboration support agents that are triggered by this analysis.

2. Conducting a series of investigations into the causal connections between conversational processes and learning as well as the causal connection between automatic interventions and collaborative behavior across multiple settings, including lab and classroom studies as well as investigations in the on-line VMT environment. This series of controlled and naturalistic observations will culminate in a large-scale summative evaluation in the on-line VMT environment.
In addition to producing new knowledge in the research area of Computer Supported Collaborative Learning, the results of this research will permanently extend the capabilities of an existing on-line math community, making it a more valuable resource beyond the end of the proposed research funding.

### 2. Foundational Resources Provided by the CMU and Drexel Teams

The CMU and Drexel teams both bring a rich storehouse of resources to the table to make use of in this effort.

#### 2.1 Technological Foundation

For a technological foundation, the CMU team brings to the project much prior work developing and evaluating tutorial dialogue technology that can be used to deliver interactive support (Rosé et al., 2001; Gweon et al., 2005; Rosé et al., in press; Rosé et al., 2005; Kumar et al., 2006; Wang et al., 2006), prior work developing automatic collaborative learning process analysis technology that can be used to trigger interventions (Donmez et al., 2005; Wang et al., 2007c), other language technologies research related to text classification (Rosé et al., 2003; Rosé et al., 2005-b), robust analysis of explanations (Rosé, 2000; Rosé et al., 2002; Rosé & VanLehn, 2005) and dialogue analysis more generally (Rosé et al., 1995; Arguello & Rosé, 2006), as well as early work on design and evaluation of adaptive collaborative learning support (Gweon et al., 2006; Wang et al., 2007; Kumar et al., 2007) and investigations of group composition and gender effects in collaborative learning in an intelligent tutoring environment (Gweon et al., 2005b; Gweon et al., 2007).

The Drexel team brings the existing Virtual Math Teams (VMT) environment (http://mathforum.org/vmt). The Virtual Math Teams (VMT) project within the Math Forum uses peer collaboration in small student teams to enhance learning and participation in math discourse. Small groups of students are invited to chat rooms (see description of the Collaborative Environment in Section 3.1) where they discuss carefully designed math problems or math micro-worlds. VMT mentors are typically not present in the chat rooms, but they provide asynchronous feedback to the student groups upon request. We proposed to augment this environment with automatic, dynamic collaboration support. Math Forum and VMT staff will be involved at all stages of designing, developing, implementing, testing,
refining and deploying these computer-support tools in close collaboration with researchers from Carnegie Mellon University. VMT researchers have extensive experience exploring the effectiveness of these materials for stimulating productive collaborative learning interactions. For analysis of collaborative discussions, VMT researchers have used a variety of methods that we will draw upon in our proposed work for on-line and off-line analysis of the learning and collaboration that takes place in the VMT-Chat environment, including statistical analysis of coded chats, ethnographic observation of participation and interaction analysis (adapting ethnomethodologically-informed conversation analysis to textual chat). A large number of studies of VMT chats are already available, including (Cakir et al., 2005; Sarmiento, Trausan-Matu, & Stahl, 2005; Stahl, 2006a, 2006b, 2006c, 2006d, 2006e; Strijbos & Stahl, 2005; Wessner et al., 2006; Zemel, Xhafa, & Cakir, 2005); see http://www.mathforum.org/vmt/researchers/publications.html for a more complete list.

2.2 Math Forum Materials

![VMT Spring Fest](image)

Here are the first few examples of a particular pattern or sequence, which is made using sticks to form connected squares:

1. Draw the pattern for \( N=4, N=5, \) and \( N=6 \) on the whiteboard. Discuss as a group: How does the graphical pattern grow?
2. Fill in the cells of the table for sticks and squares in rows \( N=4, N=5, \) and \( N=6. \) Once you agree on these results, post them on the VMT Wiki.
3. Can your group see a pattern of growth for the number of sticks and squares? When you are ready, post your ideas about the pattern of growth on the VMT Wiki.

Figure 1. Example Math Forum Problem: The Sticks Problem

Selecting appropriate materials to stimulate productive collaborative conversations is essential to fostering the success of collaborative learning. Since the goal of much collaborative learning is to stimulate higher order thinking, typical tasks used in studies of collaborative learning are open ended problems with multiple possible solutions, especially ones with many trade-offs rather than right versus wrong solutions, or highly interpretative problems such as case study analysis. We draw from resources designed by The Math Forum, which has been providing a successful, highly popular online community and digital library for K-12 students, teachers and others for over a decade (Renninger & Shumar, 2002). Although the Math Forum
works closely with school districts and teachers, its central focus is on providing informal learning experiences, by developing challenging, non-traditional math problems for students to think about and by collecting student responses. Although it has collected some of these responses into math books on algebra and geometry, it mainly organizes these responses as a digital library. In its various services (see Section 6 on Partnerships and http://mathforum.org for more details), the Math Forum facilitates interactions among students, teachers, pre-service teachers, volunteer mentors and paid staff.

An example problem is displayed in Figure 1 above. In the VMT environment, students work in small groups on the same problem over 3 sessions. In the first session, they work out solutions to the problem. In between the first and second sessions, students receive feedback on their solutions. In the second session, students discuss the feedback they received on their respective solutions and step carefully through alternative correct solutions. In that session and the subsequent session, they also discuss additional possible ways of looking at the problem including variations on that problem in order to take a step back and learn larger mathematics principles that apply to classes of problems rather than individual problems. Although the problem provides the opportunity to investigate multiple possible solutions and to engage in deep mathematical reasoning, our finding from analysis of chat logs where students have worked on this and other problems is that students tend to jump to finding one solution that works rather than taking the opportunity to search for alternative solutions. The moderator plays an important role in stimulating conversation between students, encouraging knowledge sharing and probing beyond a single acceptable solution. Thus, we plan to model our dynamic support agents after successful group moderators using a similar data driven process that was used to develop our CycleTalk tutorial dialogue agents (Rosé et al., in press; Kumar et al., 2006), patterned after successful human tutors (Rosé et al., 2005) supporting learning in the same environment that the chat agents now participate in. Examples of the proposed support are given in Section 3 below.

### 2.3 Tools for Building Dynamic Collaborative Learning Support

What the CMU team brings in terms of technological infrastructure are tools for automatic collaborative learning process analysis to trigger dynamic support in the midst of ongoing collaboration and tools for quick authoring of conversational agents to administer the interactive support. Note that both of these tool sets were developed under the NSF funded Pittsburgh Science
of Learning Center (PSLC) as enabling technology projects. Whereas in the PSLC this work can support classroom studies in designated LearnLab courses (which do not include any courses using Math Forum materials), that center does not fund work in on-line learning communities, classroom studies in other classrooms, or lab studies. Thus, the proposed work will take resources developed in one NSF funded context, and extend the impact to a new and significantly broader context.

As part of a collaboration with the Knowledge Media Research Center in Tuebingen, Germany, we have developed a proof of concept for fully automatic collaborative learning process analysis (Donmez et al., 2005; Wang et al., 2007-b; Rosé et al., in press). We describe this work here as an example of the types of analyses and level of detail we are able to achieve with our automatic processing of conversational data. We refer to the coding scheme used in this work, which was developed by Weinberger & Fischer (2006), as the Weinberger and Fischer coding scheme. This coding scheme was developed for the purpose of addressing the question of how computer-supported collaboration scripts could foster argumentative knowledge construction in online discussions. Argumentative knowledge construction is based on the idea that learners acquire knowledge through argumentation with one or more learning partners, by better elaborating the learning material and by mutually refining ideas. Argumentative knowledge construction must be evaluated on multiple process dimensions. Thus, the Weinberger and Fischer coding scheme has five process dimensions. These dimensions are derived from different theoretical approaches and focus on specific conceptualizations of argumentative knowledge construction, and are supposed to be independent from each other. The main concepts are (1) epistemic activity, formal quality regarding argumentation, which differentiates in the (2) micro-level of argumentation and the (3) macro-level of argumentation, and (4) social modes of interaction. Independent of these dimensions, the segments have been coded whether they were or were not (5) a reaction to a previous contribution.

Each dimension offers a different perspective on the nature of the contribution, often drawing upon information of a different nature from the other dimensions, and thus offers evidence of the generality of our approach. For example, the Micro and Macro dimensions each characterize different aspects of the linguistic structure of the contributions whereas the Social Modes and Reaction dimensions focus on different types of social conventions and relational styles conveyed in and encoded in contributions. Automatic application of coding schemes such as this one make it possible to automatically detect dysfunctional communication patterns within running discourse. For example, they make it possible to determine whether
participants are acknowledging each other’s contributions, and considering them adequately without either giving in too quickly or rejecting each other’s views out of hand. A major focus of our work has been increasing classification accuracy on low frequency events, since many times very infrequent events are nevertheless important to recognize with a high degree of accuracy because they are indicative of particular types of trouble.

The second technology provided by the CMU team is an infrastructure called TuTalk to support quick authoring of dialogue agents (Gweon et al., 2005; Jordan et al., 2007; Cui et al., to appear). This work includes 1) tools for non-technical users to author dialogue specifications for particular student exercises and 2) a backend system for supporting full spoken or text-based dialogue behavior that follows the authored specifications. In our prior work we have explored strategies for supporting the development of language understanding interfaces by non-linguists (Rosé & Hall, 2004; Rosé, Pai, & Argeullo, 2005). TuTalk provides a suite of corpus organizational tools to help authors prepare their corpus data in preparation for authoring using what we refer to as the InfoMagnets interface (Arguello & Rosé, 2006b). The TuTalk authoring interface is then used for finer grained processing, such as shifting topic segment boundaries and labeling more detailed utterance functionality, as well as authoring templates used for generating dialogue behavior. These tools were used to build the dialogue agents used in the successful classroom studies reported in (Kumar et al., 2006) in one week.

While our previous work developing dialogue agents has focused primarily on tutorial dialogue for individual learning, here we expand our scope to cover tutorial dialogue for collaborative learning, and have already seen success in that arena (Kumar et al., 2007-a; Kumar et al., 2007-b; Wang et al., 2007-a; Chaudhuri et al., to appear-a). Thus, here the purpose of the dialogue agents is not to lead one student to reflect on a past decision or come to a specific conclusion. Instead the dialogue agents will seek to direct the interaction between students, offering instruction only as a last resort. Building on work reported in (Rosé & Torrey, 2005), we seek to build dialogue agents that are effective at eliciting elaborated explanations from students in the context of the help seeking and help giving interactions with other students in order to implement dynamic support interventions.
3. Results from SGER: The VMT-Basilica Integrated Foundation for Supporting Collaborative Learning On-Line

In this section we describe how we have integrated elements from our previous work into a common technical foundation and have developed a foundational instructional approach that we build on and extend in our proposed work.

3.1 Collaborative Environment

The Math Forum and its Virtual Math Teams Project have collaborated closely with CMU personnel under an SGER grant to design, develop, implement, and pilot test the integrated VMT-Basilica environment. This was accomplished using the VMT-Chat environment, which was made available as a test-bed for collecting data about the performance of these tools. The opensource VMT-Chat includes the VMT Lobby, where people can select chat rooms to enter, and a number of math discussion chat rooms, that each include a text chat window, a shared drawing area and a number of related tools (for a more detailed description of the environment and how it is used, see (Stahl, 2006). Since the environment is available as Open Source, (1) it was easily extended for this project and (2) the results of this project can easily be made available to other researchers.

VMT-Chat includes the VMT Lobby – where people can select chat rooms to enter (see figure 2)

– and a number of math discussion chat rooms – that each include a text chat window, a shared drawing area and a number of related tools (see figure 3).

Three types of rooms can be created in the lobby:
• Open rooms. Anyone can enter these rooms and participate in the discussion – see Figure 1, where open rooms are listed under math problems or topics.

• Restricted rooms. Only people invited by the person who created the room can enter.

• Limited rooms. People who were not originally invited can ask the person who created the room for permission to join.

Such flexibility allows the VMT service to be used in a wide range of ways and in limitless combinations and sequences:

• For instance, teams of students from the same classroom might first use the VMT environment to work together on a series of Problem of the Week (PoW) problems during class time, allowing them to become familiar with the system and build collaboration skills in a familiar social setting.

• Later they could split up and join groups with students from other schools to explore more open-ended mathematical situations.

• As they become more advanced users, they can create their own rooms and invite friends or the public to discuss topics that they themselves propose.

VMT-Basilica integrates the open source VMT-Chat with Basilica (Kumar et al., submitted-a; Kumar et al., submitted-b), which integrates all of the CMU technologies discussed above into a framework that supports rapid development of computer supported collaborative learning environments. It provides a clean software architecture where technologies such as TagHelper tools and TuTalk are encapsulated into abstract classes that can be instantiated as specific Filters and Agents respectively, where Filters listen in on the behavior occurring within an environment like VMT-Chat in order to notice specific behaviors, either in the chat or other interface widgets such as the whiteboard, and Agents exhibit behaviors, such as displaying graphics on the whiteboard, or participating actively in the chat. Within this integrated framework, we are able to trigger a variety of interventions within the VMT-Chat environment that are sensitive to what is happening in the collaboration between students. We have already successfully piloted the integrated VMT-Basilica environment in a large classroom study related to collaborative design for environmentally friendly engineering (Chaudhuri et al., to appear).
3.2 Instructional Approach

Our goal is to maximize the benefit students receive from the interactions they have with one another. Not all instructional conversation between learners is equally effective, and often requires some form of support in order to become effective (Stegmann et al., 2004; Rummel et al., 2003). State-of-the-art forms of collaboration support proactively structure collaborative learning interactions using a broad assortment of approaches. Examples of such support includes assignment of students to roles (Strijbos, 2004), provision of prompts during collaboration (Stegmann et al., 2004), design of structured interfaces including such things as buttons associated with typical “conversation openings” (Baker & Lund, 1997), instructions to guide learners to structure their collaboration (Webb & Frivar, 1999), or even various forms of collaboration training (Rummel et al., 2006). These approaches to structuring collaboration play a role similar to training wheels on bicycles. Just as training wheels allow kids to have the experience of riding a bike before they are ready to do it independently, these forms of collaborative learning support increase the amount of productive collaboration behavior above that of what it would be without the structuring, thus allowing students to collaborate at a higher level than their own collaborative skills would naturally allow. As is well known, however, training wheels must eventually come off. And typically, they are removed by a watchful parent, who may decide after watching their child fall a few times, to put them back on for a time until the child has developed further in their own coordination and balance. In a similar vein, the learning sciences literature tells us that scaffolding should be faded over time (Collins et al., 1991), that over-scripting is detrimental to collaboration (Stegmann et al., 2004), and unnecessary support is demotivating (Dillenbourg, 2002). However, in order to fade collaboration scaffolding as a watchful parent, we must do so using technology that is sensitive to collaborative behavior in the environment. Thus, a major goal of our research is to support collaboration in a way that is responsive to what is happening in the collaboration rather than behaving in a “one size fits all” fashion, which is the case with state-of-the-art static forms of support.

Our instructional approach is modeled after constructivist principles of classroom discourse, such as those advocated in (Chapin, O’Connor, & Anderson, 2003). Webb and colleagues present a series of studies in different educational settings that demonstrate the importance of the depth of instructional explanations, both for the speaker as well as the recipient (Webb, 1991; Webb, Nemer, & Zuniga, 2002). Much research shows the value of drawing out student reasoning in the form of elaborated explanations. In particular, one of the best substantiated educational findings
in cognitive science research related to education is the educational benefit of explanation, and in particular, the self-explanation effect (Chi et al., 1989; Chi et al., 1994; Chi, 2000). Nevertheless, previous discourse analyses of collaborative conversations reveal that the majority of conversational interactions between students do not display the “higher order thinking” that collaborative learning is meant to elicit (Webb & Mastergoerge, 2003; Webb, Nemer, & Zuniga, 2002), and we have found this as well in our own observations of collaborative learning, both at the college level (Gweon et al., 2006) and at the middle school level (Gweon et al., 2007).

To begin to move past the traditional one-size-fits-all non-adaptive approaches to collaboration support, we have conducted a series of studies in which we experimentally investigate foundational issues related to the design of dynamic support for on-line collaborative learning (Gweon et al., 2006). These initial investigations demonstrated that explanation elicitation prompts such as those displayed in Table 1, delivered strategically, based on simple rules related to timing of contributions and distribution of labor between student, were effective for eliciting explanation attempts as well as significantly increasing learning. These very general purpose prompts were effective across a wide range of collaborative problem solving exercises. In our long term plans in the VMT context, in order to elicit the type of collaborative behavior that leads to more learning, we will use dynamic collaboration support agents based loosely on the style of our previous investigations at the secondary and post-secondary level (Gweon et al., 2006; Wang et al., 2007; Kumar et al., 2007).
Our previous success with automating collaborative learning process analysis (Donmez et al., 2005; Rosé et al., in press) offers promise that the dynamic support mechanism evaluated using a Wizard-of-Oz setup in (Gweon et al., 2006) can be implemented and deployed fully automatically.

We have run a number of successful pilot studies in which we used dialogue agents to deliver interactive support when triggered by an automatic analysis of the collaborative learning discussions as they unfolded (Wang et al., 2007; Kumar et al., 2007-a). In these successful studies, the fully automatic interactive support lead to significant increases in learning in comparison to a control condition that did not have the interactive support. However, these studies did not take place in an open web environment such as the Virtual Math Teams environment. Thus, there is still much work to do to investigate how best to elevate the level of helping behavior in an environment such as the on-line VMT environment.

<table>
<thead>
<tr>
<th>Case</th>
<th>Prompts</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The other student would benefit from more explanation. Please elaborate on your correction.</td>
</tr>
<tr>
<td>2</td>
<td>Help the student understand your correction. The other student seems to be struggling with this section of the problem. Please offer your assistance.</td>
</tr>
<tr>
<td>3</td>
<td>Please be sure you are working with the other student to solve the problem. It seems like the other student has not contributed lately. Why don’t you see if they need help?</td>
</tr>
<tr>
<td>4</td>
<td>It seems like you are moving on before understanding your errors. Please spend more time reviewing this page. Does the other student understand the errors made on this problem? Please share your understanding of this page with the other student.</td>
</tr>
</tbody>
</table>
4. Full-Circle Methodology: Exploring the Design Space in Complementary Contexts

We propose to take advantage of the complementary insights we can gain from investigations in various settings, including lab studies, classroom studies, and studies in the Virtual Math Teams environment. Furthermore, we leverage a broad spectrum of methodologies, ranging from high internal validity studies in the lab and in the classroom, with pre/post test designs to high external validity investigations in the “wild” Virtual Math Teams environment where the same analyses of observable collaborative behavior are possible even with naturalistic, non-controlled observation, but experimental designs are less practical and must be administered with caution because of the way imposing too much control may interfere with the natural working of the community (In Section 4.4, we describe how we will carefully conduct a large-scale summative evaluation at the end of the project in such a way as to avoid interfering with the natural workings of the community any more than necessary.) With respect to analysis of log data, we will also employ a diversity of approaches including formal, quantitative analyses of log data based on categorical coding as well as ethnographic style analyses.

4.1 Illustration of Methodology

As an illustration of our full-circle, mixed-methods approach, we offer an example of how our informal collaboration to date is already yielding synergistic findings. Because our ultimate goal is to achieve success in the “wild” Virtual Math Teams environment, we begin with insights gained from an ethnomethodological analysis of chat logs collected in the Virtual Math Teams environment (Stahl, 2006). In one notable chat session, we observed a group of students that was successful at solving problems collaboratively that none of them were capable of solving alone. On close inspection of the chat logs, a student who at first appeared as “the class clown” emerged as a tone setter in the analysis, putting his team mates at ease, and allowing them to forage ahead as a group towards solutions to very challenging problems. From this analysis, a hypothesis emerges that interventions that break the tension and stimulate a light-hearted atmosphere in a collaborative learning setting may act as a “social lubricant”, and thus may increase success in collaborative problem solving. The Carnegie Mellon team has tested this hypothesis experimentally in a classroom study, referred to as the Social Prompts study (Kumar et al., 2007-b), in which students worked in pairs in a collaborative problem solving environment
that shares some common simple functionality with the Virtual Math Teams environment.

In the experimental condition of the Social Prompts study, before a problem is displayed in the shared problem solving space, a tutor agent first asks each student what we are referring to as a social question. The idea is to draw the students into a productive, collaborative attitude by encouraging them to put something they identify with personally into the math problems, so they feel as though they have worked together to make the math problems they will then work together to solve. As an example, consider the following scenario: The agent may first ask student 1 “Student 1, if you had to choose between a long flight or a long car ride, which seems more uncomfortable?” The student indicates that a car ride would be preferable. Then the tutor agent may then ask, “Student 2, which are more entertaining – books or movies?”, and the student may respond that books are more amusing. These two pieces of information are then used to fill in slots in a template that is then used to generate the math problem. In particular, the resulting story problem says, “Jan packed several books to amuse herself on a long car ride to visit her grandma. After 1/5 of the trip, she had already finished 6/8 of the books she brought. How many times more books should she have brought than what she packed?” The lighthearted nature of the questions was meant to inject a note of playfulness into the conversation. In order to control for content and presentation of the math content across conditions, we used exactly the same problem templates in the control condition, but rather than presenting the social questions to the students, we randomly selected answers to the social questions “behind the scenes”. Thus, students in both conditions worked through the same distribution of problems.

The results of the Social Prompts study provided some evidence in support of the hypothesis that emerged from observations in the Virtual Math Teams environment. We began our analysis by investigating the socially oriented variables measured by means of the questionnaire, specifically perceived problem solving competence of self and partner, perceived benefit, perceived help received, and perceived help provided. For perceived benefit and perceived confidence, scores were high on average (about 4 out of 5) in both conditions, with no significant difference between conditions. However, with perceived help offered as well as perceived help received, there were significant differences between conditions. Students in the experimental condition rated themselves and their partner significantly higher on offering help than in the control condition. Interestingly, there is more evidence of requesting help in the control condition chat logs.
However, these requests were frequently ignored. The learning gains analysis is consistent with the pattern observed on the questionnaire and offers some weak evidence in favor of the experimental condition on learning. The trend was consistently in favor of the experimental condition across tests and across units of material on the test. The strongest effect we see is on lab day 2 where students in the experimental condition gained marginally more on interpretation problems (p=.06, effect size .55 standard deviations). The student chat logs contain rich data on how the collaborative problem solving process transpired. We conducted a qualitative analysis of the conversational data recorded in the chat logs in order to illuminate the findings from the tests and questionnaire data discussed above. Overall, we observed that students were more competitive in the control condition. Insults like “looser”, “you stink”, “stupid” occurred frequently in the control condition, and never in the experimental condition. Instead, in the experimental condition we observe light hearted teasing. There were significantly more help related conversational episodes per problem in the Experimental condition (Kumar et al., 2007-b), and furthermore, it happened significantly more frequently in the Experimental condition that when students got stuck on a problem solving step, they received help and then were able to complete the step themselves instead of their partner completing it for them, which was the general case in the control condition.

The full-circle methodology that we propose begins with ethnographic observations from interactions in the Virtual Math Teams environment. These observations lead to hypotheses that can be tested in high internal validity environments such as lab and classroom studies. These studies help us to confirm causal connections between stimuli and subsequent effects, between which we observe a correlational relationship in our earlier ethnographic analyses. Discovered causal connections can then form the basis for the design of full-scale interventions that can then be prototyped and tested in the Virtual Math Teams environment. These investigations can eventually serve both as a test of the generality and robustness of findings from the lab and classroom studies as well as providing new insights that form the basis for new hypotheses that can then be tested in further cycles, although only a large-scale controlled study, as we propose for in the final year of the project, can provide definitive evidence in favor of an intervention. In our three year project, we propose three complete cycles, ending with a carefully designed, large scale experimental study in the Virtual Math Teams environment to verify the effectiveness of the interventions we will develop in that environment, as detailed in Section 5.
4.2 Investigation of Contextual Variables in the VMT Environment

From lab and classroom studies where we are able to use pre and post tests, we are able to determine which types of interactions are more conducive to learning than others. We have already conducted a series of successful classroom studies investigating questions related to the design of computer supported collaborative learning environments (Gweon et al., 2007; Wang et al., 2007; Kumar et al., 2007-a; Kumar et al., 2007-b; Chaudhuri et al., to appear; Kumar et al., submitted-a). One major question we address in the proposed naturalistic observations in the on-line VMT environment as well as the large-scale summative evaluation in that environment in year three is whether or to what extent we can use the same interventions in “the wild” to achieve the same effect on behavior that we observe in the lab or in the classroom. This behavior is directly observable from the logs we collect. Thus, we can investigate these important questions about the effect of alternative interventions on collaborative behavior in the VMT setting even without pre and post tests. Specifically, what we seek to learn from our investigations in the on-line VMT environment is how the contextual variables that distinguish that environment from the lab and the classroom environments may interfere with or change the effects of interventions on student behavior. Such variables include the time of the interaction (e.g., during school hours, in the evening, on the weekend, during the summer), location (co-located or distributed teams), reward structure (e.g., in class assignment, homework assignment, extra credit, or voluntary), group composition (e.g., same grade cohort, mixed grade/age), collaborative pre-disposition (e.g., students coming from schools where collaborative learning is encouraged and frequent versus schools where collaboration is not part of regular instruction), and experience in the environment (e.g., new to the on-line VMT environment versus having participated for a long time). We will carefully keep track of this information about students and take them into consideration as we interpret findings from naturalistic observations used for hypothesis formation. In order to test these hypotheses some of these variables will be manipulated in a quasi-experimental manner in the large-scale summative evaluation in year three.

4.3 Experimental Paradigm

All lab and classroom studies will use the following experimental paradigm.

Participants. Participants will be middle school children recruited through local newspapers or through their teacher and will be randomly assigned to pairs, which will then randomly be assigned to conditions. We recognize that many characteristics of students may interact with our experimental
manipulations such as ability level of individual students, differences in ability level of students in pairs, gender of individual students as well as gender mix of pairs, level of interest and motivation of individual students. In order to accommodate this, we will recruit at minimum 20 pairs per condition in each study in an attempt to achieve a balance of all of these factors, and we will include these variables in our analyses.

Materials. All instructional materials including tests, questionnaires, and problem solving activities will be based on existing Math Forum materials, and will be adapted for studies by researchers both at CMU and the Math Forum working in close collaboration. We will also seek guidance from the math coach who is our partner at Propel Charter School (See Section 6). All collaborative work sessions except for controlled studies in year one will take place in the VMT environment described in Section 3.1.

Experimental Procedure. We will strictly control for time in all experimental studies. Each pair will participate in a single two hour session, which includes time for pre and post tests, in some studies a supplementary tutoring session, and group work. In all cases, the experimental manipulation will take place during the group work segment. In studies with a supplementary tutoring segment prior to group work, students will also take a middle test prior to group work in order to separate learning from tutoring from learning during group work. Pre, post, and middle tests will be isomorphic, and we will counterbalance the order of the tests in order to control for any potential differences in difficulty between tests, as in our prior work (Gweon et al., 2006; Gweon et al., 2007). As in our previous studies, students will also take a questionnaire at the end of their participation to assess their perceptions of the collaboration, their attitudes toward their mathematical learning and the on-line learning environment overall.

Experimental Manipulation. Based on our previous experience, with 20 pairs per condition, we expect each lab study to require 6 weeks times the number of conditions. Thus, a 4 condition study would require about 6 months to run. Allowing time for analysis of results and reflection in between experiments, we expect to run between 4 and 6 studies of this magnitude, or fewer larger studies, within the 3 years of the proposed work. Each study will include a control condition with fully unsupported collaboration in order to obtain an accurate measure of the value of each intervention. Some experimental manipulations, such as ones involving choices about which resources to provide students with, do not require sensitivity to the ensuing collaborative process, whereas others require detecting patterns of collaborative behavior that are indicative of trouble in
the collaboration. In early lab studies, as we are continuing to extend the capabilities of our automatic process analysis technology to the specific demands of our proposed work, interventions will be triggered using a Wizard-of-Oz setup as in (Gweon et al., 2006; Benzmueller et al., 2003), where an experimenter is watching the collaboration remotely and selecting interventions at key points in the process. As the technology becomes reliable enough, we will replace the human intervention with automatic triggering of interventions.

Process Analyses. As in our prior studies of collaborative learning, in addition to analyses of test and questionnaire data, we will explore the collaborative process through analysis of the chat logs collected during group work (Meier et al., submitted; Weinberger & Fischer, 2006; Strijbos, 2004; Lally & De Laat, 2002). Variables related to group process such as amount of deep explanation behavior, help seeking and help provision behavior will be analyzed both as ends in themselves, i.e., examining the effect of our experimental manipulations on patterns of communication, but also as mediating variables in our comparisons of pre to post test gains and questionnaire findings.

Prior to each formal study, we will run several pilot testing sessions for each new condition in order to fine-tune our execution of our experimental manipulation.

4.4 Example study: Eliciting Helping Behavior with Dynamic Prompts

In our previous investigations with middle school students, we have observed that one area of needed support in collaborative problem solving is supporting the generation of explanations. We can offer some non-interactive support for this by means of fully worked out examples that include explanations, glossaries that define technical terms required for understanding the problems, and examples of clear explanations contrasted with unclear explanations. This type of support is in the spirit of the type of non-adaptive support for collaboration offered to students in state-of-the-art collaborative learning environments. What we propose to investigate that builds upon this prior work is the interaction between these non-adaptive forms of support and an adaptive form of support similar to the prompts used in our prior exploration of adaptive prompting with college aged students (Gweon et al., 2006). In that study, the adaptive support we offered students instructed them when to offer help but not how to offer help. But with middle school students (Gweon et al., 2007), we observed that students sometimes realized they should offer help but were not able to.
Thus, our first lab study we will contrast Non-adaptive support versus No support versus Non-Adaptive support versus Adaptive support (in the form of simple adaptive prompting as in (Gweon et al., 2006)). The purpose of the tutorial dialogue agents will be to scaffold the process of constructing an explanation by drawing the explanation out of a student step by step. We hypothesize that students will be better equipped to offer help in the Non-adaptive support condition than in the No Non-adaptive support condition, but may not be significantly more likely to attempt to offer help unless they have the additional support of the Dynamic support agents. We further hypothesize that low ability students will benefit more from the more elaborate form of Adaptive support than high ability students.

Here’s an example knowledge construction dialogue agent, which can be used to deliver concept specific help. This is an example of the type of dialogue agent that can be authored using our development tools in under an hour (Cui et al., to appear; Jordan et al., 2007; Gweon et al., 2005-b), and which could be used for this study as a form of support for collaborative problem solving. Note that both students have the opportunity to participate in the conversation, and the agent offers students the opportunity to practice explaining math concepts with coaching. The goal of the conceptual dialogues is to encourage students to think about the ideas behind the fraction arithmetic problems.

*Tutor*: Let’s think about why the flip and multiply strategy works. Consider this: If I have 1 divided by 1/2, can I turn that into something divided by 1?. If I want to turn the denominator, which is 1/2, into 1, what do I need to do to it?

*Student1*: I don’t know how.

*Tutor*: You need to multiply by two. And after you do that, what do I have to do to the numerator?

*Student1*: also multiply by 2

*Tutor*: Right. When you multiply the numerator and denominator of the problem by the denominator of the fraction that is the denominator of the problem, you see a simpler version of the problem. So why does flipping and multiplying work?

*Student2*: what you’re doing is multiplying the numerator and denominator by the same number

*Student1*: ya, multiplying by the denominator of the bottom fraction

*Tutor*: Right. So let’s try another one. If I want to divide something by 2/3, what do I need to multiply the numerator and denominator by?

*Student1*: 2/3

*Tutor*: Right. You should continue to think about why it's 3/2.
We estimate that we will need to author approximately 6-10 such dialogue agents for a weeklong study. Since the agents target general mathematic concepts rather than specific errors or specific math problems, they can be used in a variety of different contexts.

4.3 Subsequent Lab Studies

The series of studies that we run under this grant will build one on top of the other in terms of results. Thus, it is not possible to fully plan out the exact series of studies that we will run as we fine tune the design of our collaboration support approach and accumulate findings from study to study. However, we have specific ideas about alternative follow-up studies planned after the initial one just mentioned. For example, one question is whether our adaptive support should emphasize encouraging help providing behavior or help seeking behavior. Students may be more motivated to respond to a help request coming directly from their partner student rather than a request to offer help to that student when the request comes from a computer agent. Furthermore, students may be more aware of the specifics of the need for help when the help request comes directly from the student who needs the help, and thus the manipulation of prompting help seeking behavior versus prompting help providing behavior may have an effect on the quality and specificity of the help that is offered. A similar manipulation would be contrasting prompts that simply request that help be offered to the partner student versus prompts that refer to specific types of help or help on a specific topic when that request comes from a computer agent.

4.4 Evaluations in the VMT Environment

The computer-based tools developed under the proposed grant will be tested in naturalistic observations in the on-line VMT environment on a small scale throughout the 3-year project, and will be evaluated in a large-scale summative evaluation in the 3rd year of the project.

The tools will be used in five ways:

1. In early naturalistic trial cases in the VMT environment, rather than directly intervening in student collaboration, instead the assessment of the collaborative learning interactions provided by the automatic process analysis technology discussed in Sections 2.3 and 3.3 will be provided asynchronously to human mentors who provide feedback to students between student sessions.
2. In a few trial cases, mentors will be in the chat room while the students are interacting. The mentors will use real-time data from the tools to provide synchronous mentoring to the students.

3. As the tools become more reliable, the support agents will interact with students within the environment, but in a mode where human moderators can intercept the messages when necessary.

4. When the agents have reached an acceptable level of performance, real-time support from the tools in the style found most successful in our lab studies will be provided synchronously to the students themselves during collaboration.

5. In all cases, explorations in the VMT environment will be more naturalistic than in the lab and classroom settings. Analysis of the naturalistic trial cases will mainly take the form of case studies. In the small scale evaluations in the VMT environment in the initial segment of the research project, brief interactions will be analyzed in detail to assess the impact of the data from the tools. Investigations in on-line settings cannot as easily be controlled and replicated to meet the requirements of traditional quantitative analysis. Therefore, qualitative interaction analysis is generally used in design-based research where conditions are changing as technology is redesigned and as the understanding of human participants also evolves (Design-Based Research Collective, 2003; Hutchins, 1996; Koschmann, Stahl, & Zemel, 2006; Maxwell, 2004). We expect these observations to complement the more quantitative findings from our controlled investigations. Their value comes from the highly externally valid insights we will gain.

As a final acid test of the technology, in the final year we will run a large-scale evaluation in the VMT environment. We will endeavor to conduct this evaluation under as realistic of circumstances, true to how the VMT environment typically operates, as possible while maintaining enough experimental rigor to obtain generalizable and robust results. We will recruit students in the same way that students are typically recruited to participate in the VMT environment. Students who agree to participate will be given a pretest to assess their level of competence with the subject material going in to the study. We will ensure that this VMT “sub-community” does not mix with the larger VMT community during the time of the study, but beyond that we will not dictate the frequency or timing of their interactions in the environment any differently than typical VMT students. More specifically, there will be two such “sub-communities” for this study. In the control condition “sub-community”, students will only
receive the support that is currently offered in the VMT environment, specifically where limited support is offered by human moderators asynchronously. In the experimental condition, students will receive this support in addition to support by fully automatic support agents that will participate in all of their on-line interactions in the VMT environment for the duration of the study. We will keep careful track of when and how long each participant is logged into the VMT environment so that we can take this into account in our analysis. At the end of the study, students will take a post-test. We compare conditions in terms of (1) pre to post test learning gains, (2) time on task, and (3) amount of observed helping behavior.

5. Research Plan Overview Integrating Research and Education

Rosé will oversee all work conducted at CMU, which includes basic research on automatic collaborative process analysis and interactive collaboration support delivery as well as lab and classroom studies. Stahl and Weimar will oversee all work conducted at Drexel University, which includes ongoing development of the VMT environment and conducting naturalistic observations on-line in the VMT environment. The CMU and Drexel teams will collaborate closely to design studies that will take place in the VMT environment, as well as analyzing the data collected in those studies, culminating in a large summative evaluation in the VMT environment in the final year of the project. The CMU and Drexel team will conduct phone conferences twice a month to coordinate their efforts. The timeline of the proposed work will be organized around three cycles of the methodology proposed in Section 4:

**Year 1.** From the beginning of the project, all lab and classroom studies as well as the naturalistic VMT environment observations will be conducted using VMT-Basilica developed during our existing collaboration. Early design experiments in the on-line VMT environment will use a hybrid methodology where the behavior of an automated agent is enhanced by the involvement of a human behind the agent as in (Rosé & Torrey, 2005). At the same time, we will conduct the lab study proposed above in Section 4.4 using the existing VMT-Basilica framework. The study described in Section 4.4 already builds on our prior results and observations, and thus is consistent with our proposed mixed methods methodology. In the second half of year 1, we will elaborate the VMT-Basilica framework based on findings, as well as using this analysis to plan the next cycle of experimentation.
Year 2. During Year 2, in addition to running the next cycle of lab/classroom studies and observations in the VMT on-line environment, we will continue to extend the capabilities of our automatic collaborative learning process analysis technology in directions motivated by findings from earlier cycles of research.

Year 3. The final year of the project will proceed as Year 2 except that in the final 6 months of the project we will conduct a large-scale summative evaluation study in the VMT on-line environment, as proposed in Section 4.4.

PIs Rosé and Stahl both teach courses in Computer Supported Collaborative Learning, which under this grant will be integrated into a single distributed course. Both courses involve a significant project component, in which distributed teams of students from both universities will join forces to participate in the research. One such opportunity they will have will be to prototype dynamic collaborative learning support interventions using the tools provided by the Carnegie Mellon team, which will then be pilot tested in Drexel’s Virtual Math Teams environment. Analysis of chat logs from Virtual Math Teams interactions, especially involving dynamic support agents, will also be a course activity. Thus, students in the courses will not only benefit by learning about the findings from the research, but they will also actively participate in the research. Whereas the course at Drexel emphasizes a socio-cultural approach to computer supported collaborative learning, the course at Carnegie Mellon has more of a cognitive emphasis. Thus, the distributed teams will provide an ideal environment for wrestling with issues on the frontier between these two communities and gaining greater insight into the deep connections between the social and cognitive foundations of collaborative learning.

Results from the proposed research will be presented in conferences and journals in the fields of computer supported collaborative learning, human-computer interaction, and computational linguistics.

6. Partnerships

We have an ongoing partnership with Propel Charter School in Homestead, Pennsylvania where we have run a Math Camp in summer of 2006 and subsequently ran an after school math club in order to collect data on math explanations and collaborative behavior from urban middle school kids in connection with the specific Math Forum materials we have used and plan to continue use in our studies. This partnership provides one context where
the CMU team will do outreach work using the on-line environment that provides support to students outside the classroom as part of this project. See letter of support from Propel Charter School’s math coach, Ariane Watson, written in support of an earlier proposal related to collaborative math problem solving that was not funded but nevertheless served as a stimulus for beginning to build this partnership in anticipation of an eventual funded research project.

The Math Forum at Drexel University, run by Co-PI Steve Weimer, manages a website (http://mathforum.org) with over a million pages of resources related to mathematics for middle school and high school students. This resource actively provides mathematics instruction to the full gamut of American students, but specifically targeting a very significant segment of low income and minority students. A leading online resource for improving math learning, teaching and communication since 1992, it is now visited by over a million different visitors a month. A community has grown up around this site, including teachers, mathematicians, researchers, students and parents using the power of the Web to learn math and improve math education. Studies of site usage show that students have fun and learn a lot; that educators share ideas and acquire new skills; and that participants become more engaged over time.

7. Results from Prior NSF Funding

Rosé has supervised NSF DRL-REESE/SGER-0723580 (Exploring Adaptive Support for Virtual Math Teams, $49,999.00, July 2007 - June 2008). This SGER project provided strategic funds to develop and pilot test the VMT-Basilica environment described earlier in this proposal (Cui et al., in press; Kumar et al., submitted-a; Kumar et al., submitted-b; Chaudhuri et al., to appear). It builds on Rosé’s earlier research supported by NSF EHR/SGER-0411483 (REC: Calculategy: Exploring the Impact of Tutorial Dialogue Strategy in Shaping Student Behavior in Effective Tutorial Dialogue for Calculus). This SGER project provided the foundational research towards the concept of adaptive collaboration support that this proposal is built upon (Gweon et al., 2005; Gweon et al., 2006; Kumar et al., 2007a; Kumar et al., 2007b; Wang et al., 2007). The most successful of these early studies (Kumar et al., 2007a) demonstrated that students working in pairs with adaptive support provided by tutorial dialogue agents learn 1.24 standard deviations more than students working alone without this support. Other publications from this work include foundational work for the subsequent TagHelper tools project (Gweon et al., 2005b), with
subsequent work and downloadable toolkit at http://www.cs.cmu.edu/~cprose/TagHelper.html, which has a user base of over 350 users in over 30 countries.

CoPIs Weimar and Stahl have jointly supervised the Virtual Math Teams (VMT) project at Drexel University. NSF DUE 0333493 Collaboration Services, $450,000, August 2003 to July 2005, NSF REC 0325447 Catalyzing & Nurturing Online Workgroups, $2,299,978, September 2003 to August 2008. Virtual Math Teams (VMT), led by Gerry Stahl, Drexel University, College of Information Science and Technology, Steve Weimar, Director of The Math Forum @ Drexel, and Wes Shumar, Associate Professor, Culture and Communication, Drexel University: The VMT Project investigates issues of online collaborative mathematics problem solving by extending the Math Forum’s popular “problem of the week” service for use by small groups of students. These issues include the pedagogy of online collaborative learning of school mathematics, the design of appropriate software and the methodology of empirical research in such settings. The VMT project has produced about 80 publications in journals, conferences and books (http://vmt.mathforum.org/vmt/researchers/publications.html). Six PhD dissertations are underway analyzing project data.

References


Treisman, P. M. (1985). A Study of the Mathematics Performance of Black Students at the University of California, Berkeley. PhD Dissertation from the Graduate Division of the University of California at Berkeley, program of Science/Mathematics Education.


SLC: Engaged Learning in Online Communities

The Catalyst project will plan a Sciences of Learning Center (SLC) focused on Engaged Learning in Online Communities and develop an interdisciplinary network of researchers to design a rigorous research agenda for understanding what online engaged learning is and could be at the individual, small group and community levels.

Vision. The richness of interactions fostered by the Web and efforts to leverage that potential richness like the NSF’s NSDL program have spawned numerous innovative spaces for learning in small and large collaborative groups within informal and formal contexts. Such approaches may overcome barriers to learning based on geographic location, time constraints, gender, initial interest, self-confidence, minority status, age, disability, or skill levels of learners.

Background. The complexity of interacting factors and the differences among the individual, small group, and community dimensions of online learning have not yet been well defined or systematically studied. While individuals’ interest has been found to gate attention, goal-setting, and learning strategies of learners in the physical world, little is understood about what takes place when deep and engaged learning occurs in online communities, how group configurations and community structures matter, or how learning by online groups can be supported to develop and be sustained over time. Experiments to date have primarily been descriptive formative evaluation studies, often focused on the individual unit and situated in particular social contexts that are not necessarily generalizable. The proposed project will review what presently is understood about learning dynamics in online communities, including hypotheses about the location of knowledge produced and the ways in which learning interacts and evolves for individuals and for groups in the evolution of online communities. Research will target (a) the cognitive and affective relation between learners and the groups or communities in which they participate and (b) the forms of joint-activity that learners engage in online, including the psychological and social characteristics of these activities.
**Catalyst Approach.** This Catalyst project will bring together established researchers from the US and abroad who have already begun to explore engaged learning of mathematics and science online. An interdisciplinary group of lead researchers will work with the PIs to develop the SLC research agenda and proposal through small group work online and off. The work of the PIs and lead researchers will be closely coordinated with activities of international research networks. A conference for invited researchers will refine and expand the scope of research conducted by the several focused workgroups. A journal special issue will motivate and report on the resulting research agenda to be pursued by the proposed SLC.

**Intellectual Merit:** The proposed Catalyst brings together a critical mass of researchers from a spectrum of approaches who have already made significant contributions to this timely area of the science of learning. The project will identify detailed and rigorous methods to study the relations between cognitive and affective components of online learning at three levels of analysis: the individual, small group, and community. Such knowledge can help design new global learning opportunities, regardless of minority status, age, disability, self confidence, initial interest, or skill levels.

**Broader Impacts:** The world of the 21st century will involve technologically mediated spaces, where online life-long learning will take place. Work-spaces, museums, schools, healthcare facilities, and other social institutions increasingly combine physical and virtual interactions. The work of this Catalyst and the SLC research agenda that it will define are essential to a world that increasingly involves computer-supported cooperative work and collaborative learning. The principles developed for mathematical and scientific learning can be explored and adjusted for other disciplines and new forms of formal and informal learning within online communities.

**Vision**

The world is increasingly mediated by advanced communication technologies and all social spaces are becoming hybrid spaces combining traditional physical space with social cyberspace. This is evident in the ways that cell phones, handheld computers, and the Internet are used. It also is clear in the ways that work-places and businesses are being reorganized around the flow of information. The most economically developed societies have made such significant social investments in communication and information technology that the sociologist Manual Castells (1996,1999) has argued for what he calls “the information society”.

Online communities appear to have a tremendous potential to reach a mass audience and to support engaged learning. Lenhart, Rainie, and Lewis (2001) reported that 73% of youth ages 12 through 17 use the Internet (surely even higher in 2005). Almost all online teens (94%) use the Internet to do research for school. The proportion of online girls that has used instant messaging (IM) exceeds that of boys -- 78% for girls and 71% for boys. Also, girls begin IM at an earlier age, with 72% of girls 12 to 14 using the service, compared to 60% of boys the same age who use IM.

New technologies have helped to form fluid linkages (e.g., between work and school) where there used to be abrupt divides. These technologies form rich socio-technical networks that have come to constitute life in this digital age, and participation in these networks is becoming commonplace. They exist in various stages, forms and venues, in non-profit Internet forums, newsgroups, and successful online professional learning communities such as the Math Forum (mathforum.org), Digital Library for Earth System Education (DLESE; dlese.org), and Hawai‘i Networked Learning Communities (HNLC; hnlc.org). They also are evident in diverse and highly profitable enterprises such as multi-player video games, online courses, and consumer services (Match.com, Amazon.com, eBay). In fact, the socio-technical world cuts across socioeconomic and international lines and public access is increasing.

Although the Internet is widely used for socializing, conducting business and information retrieval, it is rarely used for the kind of knowledge-building and deep learning that is needed in a knowledge society. In fact, the online mode tends to pressure people into quickly locating facts or registering opinions without becoming engaged enough to foster more complex understanding of a topic, such as a scientific or mathematical theme. There are aspects of the online experience, such as persistence of text and computational support that suggest an untapped potential for individual, group and organizational cognition that is rare to find in the Internet today (Stahl, in press). **How can the technical potential of online engaged learning be realized in concrete social settings?**

Online learning is clearly a new context for learning that needs to be understood, because it is increasingly being used as a context for education. Online learning spaces can be highly reflexive. There are spaces in which the learner, as well as the mentor or teacher and the researcher can look at and reflect on the process of learning. These interactions and reflections can be used as tools to support other people’s learning. Further, the creation of online spaces through cell phones, handheld computers, and the Internet can mean that learning interactions are occurring in cyberspace. This context
Proposals for Research

raises questions about the nature of the relation between computer-supported collaborative learning environments, human learning and human development: When deep and engaged learning occurs in online communities, what is taking place and how can it be supported and sustained?

Online learning contexts have characteristics that allow for studies of processes in archived data such as reflection, revision, and questioning that cannot be undertaken as easily in the physical world. They also afford multiple options for engagement—including synchronous and asynchronous communication, writing, browsing, images, video, etc.—each of which has the potential to change the way in which a person represents and understands information they work with. This context also raises questions about the relation between a person’s sense of him or herself as a learner online and the activities he or she takes on: What new forms of joint activity online are responsible for different aspects of human development, and what are the psychological and social consequences of these forms of activity?

Online learning communities and other informational technologies also allow networks to be formed quickly and easily. Often these networks can involve experts as well as novices. The communities that get formed allow for an easy kind of apprenticeship, and so readily become communities of practice (Lave & Wenger; 1991, Wenger, 1999). They have the potential to be very democratic, allowing many voices to speak, in addition to making visible the development and construction of learning to all who join the community. Understanding how collectivities get formed and under what conditions they best operate are critical issues(Klamma, Rohde & Stahl, 2004). Thus, this context also raises questions about leadership in online communities, contributors to the community, and sustainability of the community: What are the necessary conditions for the development and sustainability of online learning communities?

Understanding the interaction between the individual and contents of interest is critical. This interaction gates attention, goal-setting, and learning strategies (Hidi & Renninger, 2005). This interaction also enables learning interventions (Renninger, Sansone, & Smith, 2004). What is motivating to individuals and to the group? How are individual interests influenced by and also influencing the development of the group? In the learning sciences there has been a tendency to create a binary relation between individual and what might be termed situated cognition or group cognition. In online learning interactions, there is often a dialectical interplay between the individual and the community (Stahl, 2004). There appears to be valuable
interaction going on in some moments online where the formal and informal meet and individual interest can be reshaped as individual-group-community interaction develops (Renninger & Shumar, 2004). This context raises questions about the learning opportunities in working with online communities, specifically the relation between the cognitive, affective, and social moments in learning and the flexibility inherent in the construction of online learning contexts: **What is the relation between learner development and the online collectivities in which learners participate?**

The online environment is new and it needs to be fully understood. It is being used widely and its potential for supporting individual learners to learn through collaboration in small groups and/or in the larger context of community that is associated with sites needs to be examined carefully. Because of its flexibility and the archiving that is possible, this environment affords study of issues central to the learning sciences that have not previously been able to be studied, as well as new dimensions of these issues that the online context introduces.

**Background**

This section presents an overview of research at the main centers involved in this Catalyst project: The Math Forum (e.g. Catalyzing & Nurturing Online Workgroups to Power Virtual Learning Communities –VMT, NSF REC 0325447), The Digital Library for Earth System Education (DLESE) Program (e.g REC 0215640), The Wisconsin Center for Education Research, and The Hawaii Networked Learning Communities (e.g., NSF Rural Systemic Initiative REC 0100393). Themes from prior research at the four main centers involved in this catalyst are brought into dialogue with other related research in the Learning Sciences to frame some of the issues for the development of the proposed center.

**Online Contexts Redefine and Increase Learning Opportunities**

As part of the NSF-supported *Virtual Math Teams* (VMT) project at the Math Forum, PI Stahl and collaborating researchers Weimar and Shumar currently investigate online collaborative problem solving in mathematics while addressing issues of software support for collaborative learning services within a digital library. Research to date has focused on the study of student collaboration via synchronous online collaboration in the context of university courses and Math Forum’s Problem of the Week learning service. In addition research activities have identified key features for software
support such as enhanced mathematical communication, support for organizing the results of conversations as they emerge, and mechanisms for facilitating contributions, in terms of both group attention to all contributions made and creating a participatory space for all students. Further research questions include: (a) How is mathematics done by online small groups of students such that we can say, for instance, that the group is displaying deep mathematical understanding versus simply manipulating things algorithmically without such understanding? (b) What methods are used systematically by small groups in online, text-based environments for taking turns, keeping interaction flowing, repairing mistakes or misunderstandings, opening and closing sessions, constituting the group as a collectivity, etc.? (c) Can online events, activities and environments be designed to stimulate group cognition and to lower the barriers to participation and group success? (d) How can math discourse communities be catalyzed, grown and sustained by networks of small groups interacting with each other?

Research at The Math Forum has also contributed to the potential of online contexts to increase access to fields such as mathematics and science that have not previously been accessible to all learners. Although preliminary, studies of online learning have begun to suggest that these contexts can lead to new senses of possibility (Markus & Nurius, 1986; Shumar & Renninger, 2002), because they enable learners to explore and to shift their identities as learners (Linehan & McCarthy, 2000). Learners through online scaffolded support are better able to see themselves as competent in a particular area of study and relinquish their incompetent identity (Renninger & Shumar, 2002. It appears that the Web and online learning could be a context for supporting the development of and/or deepening of interest, and consequently the attention, goal-setting, and learning strategies that learners bring to engagement (Hidi & Renninger, 2005). Such findings further suggest that the context of learning in online communities may make a difference for learners for whom access to subjects such as mathematics and science has been challenged previously.

**Online Contexts and Theoretical Approaches to the Study of Complex Learning**

Theory and research on online learning need to focus analyses, not simply on specific components such as the web environment itself, the student behaviors and interactions, or the designs of learning activities, but rather on the complex interactions among many factors that cannot sensibly be broken
Proposals for Research

apart and studied individually. This will involve developing theoretical and methodological approaches that view learning environments as complex systems, a likely goal of our center. A complex systems approach has been foreshadowed throughout the years by Bransford et al.’s (Bransford, Brown, & Cocking, 2000) developing analyses of learning environments as learner-centered, assessment-centered, and knowledge-centered within larger social contexts. Activity theory (Engeström, 1999, 2001), in which the smallest unit of analysis is an entire complex activity system, has become an increasingly important theoretical lens for studies of learning with new media. Arrow, McGrath and Berdahl (2000) have also proposed a research agenda for the study of small groups as complex systems, with wide implications for the field of experimental social psychology.

Online learning as a complex system is explicitly addressed in the work of Derry and colleagues (Derry, in press; Derry & Hmelo-Silver, in press) who have used the activity field construct to analyze complex interactions among student, task, facilitator and tools within online learning environments they created for teacher professional development. Their work is conducted in STELLAR (Socio-Technical Environment for Learning and Learning Activity Research), a general theory-based system they have created for designing and supporting online courses. They used STELLAR to create web-based courses in cognitive science for pre-service teachers, which were offered and studied in two university settings over several years. Online activities in these courses were explicitly designed to foster transfer of course ideas to professional practice through activities that systematically integrated text and video case study with problem-based learning. This work produced: 1. A theoretical model for online instruction on a large scale that addresses a continuing major problem: the failure of most college classrooms to teach conceptual content in ways that insure transfer to professional practice; 2. Extensive online video, text materials, instructional activities, and online tools for supporting this instructional model to teach cognitive science to future teachers; and 3. Empirical support from controlled and quasi-experimental studies for the STELLAR approach and theory, as well as hypotheses for future research on online learning.

Since 2002, students’ evaluative ratings of STELLAR activities and tools have been stable and positive, and suggest students’ preferences to work collaboratively rather than individually on activities (Derry, in press; Derry, Seymour, Lee, & Siegel, 2004). Derry and Hmelo-Silver (in press) developed psychometrically validated concepts-in-use rubrics to score the quality of pre-service teachers’ discourse and products created in this environment. These data are important for understanding the level of specific acquisitions that are possible in the kinds of online environments...
that can be created with STELLAR. Stepwise regression analyses with these
data suggested relations between preservice teachers’ experiences on line
and their actual learning outcomes, as well as their perceptions about how
much they learned and their beliefs about the needs of the pupils with whom
they work. These multiple layers of information are important for
instruction, understanding student motivation and needed support, and
conceptualizing the multi-dimensional character of learning. These types of
archived data are simply unavailable except from the online environment.

In their study of online learning, STELLAR researchers employ For
example, STELLAR researchers integrates findings from Cognitive
Flexibility Theory (CFT) (Spiro, Collins, Thota, & Feltovich, 2003; Spiro,
Feltovich, & Coulson, 1992), online professional development approaches
through video case studies (e.g., Segoe, 2002), and from related cognitive
theories of case-based reasoning (Kolodner & Guzdial, 2000), embodied
perceptual learning (Glenberg, 1997), Schwartz and Bransford’s work on
reflection as a scaffold for transfer, and self-regulated learning (Azevedo,
Guthrie, & Seibert, 2004; Pintrich, 2000; Winne, 2001). Like the STELLAR
project, our center will approach online learning from multiple theoretical
perspectives.

### Engaging learners through participation

SELLLAR, Math Form and DLESE illustrate online learning environments
that are committed to engaging learners through participation in authentic
and personally relevant problem solving. Relevance and authenticity are
strong motivators and have been found to help learners make connections
between content knowledge and real world applications (Zech, Vye et al.,
1998). This is a cyclic process as Verschaffel and De Corte (1997a) point
out. However, learning in authentic contexts is not easy and goes against
the grain of traditional instruction. Instruction organized around authentic
problems may not honor the structure of the discipline being taught; problems bring concepts together in varied combinations. Thus online
environments that engage students in authentic problem solving must
provide well-designed facilitation and scaffolding to help students make
sense of problems in terms of their previous, current, and future experiences
(Salomon & Perkins, 1989). As a recent special issue of the Journal of the
Learning Sciences on scaffolding shows, there are many theoretically
important research issues associated with online scaffolding of authentic
problem solving (Pea, 2004). These issues will be addressed in the work of
our center.
Researchers and educators concerned with transfer have long used problem solving to help connect classrooms to problem contexts that are far removed from the instructional setting. However, an alternative that is often preferable to simulated problem solving is learning through engagement with the actual context of practice. Although many design issues remain to be resolved, online collaborative communities are well suited to support this kind of learning.

As an educational digital library, The Digital Library for Earth System Education (DLESE), like STELLAR and The Math Forum, places particular emphasis on supporting interactions between educators, students, and resource creators and developers, in both face-to-face and online communities (Marlino, Sumner, et al. 2001). The heterogeneous nature of the DLESE communities prompts an expansion and reconsideration of what learning may involve, moving beyond the context of the classroom, and towards a consideration of practitioners, content generators, and students considered as technologically supported communities. In particular, the expansion of technologically-supported learning contexts implies an equivalent expansion in the social and technical complexity of such contexts. In such an approach learning and knowledge can be seen not just as something passed between educators and students, but as general properties of wider social-technological networks; for instance, in a network educators may learn new practices from other educators, which they can then apply their own teaching contexts. Learning is relatively easy to track in a classroom, but how therefore is it to be conceptualized in networks of teachers, students, and resource creators – in what ways for instance does each group learn from the other groups in the network? DLESE researchers have begun an effort to answer questions such as these (Khoo 2001b, Khoo 2004, Sumner, Khoo et al., 2003).

Like Fischer (January, 2003), we argue that more online programs could encourage students at all levels to be life-long, reflective learners who employ new media to conduct research and collaborate with others to solve important problems. Emerging technologies and the new social discourses they afford enable and push us to conceptualize new systems for learning in which there are stronger linkages between learning environments and peoples’ everyday lives. Increasingly, “students” of the future will learn in the process of living, working and playing in a world where people of all ages and backgrounds participate in local and global learning communities made possible by new media. How to design and scaffold online communities that support learning in real-world contexts will likely be a major part of the research agenda for our center.
Fostering Development of Collaborative Online Communities

Almost all authentic problems transcend the individual human mind and require collaboration, since knowledge is distributed across domains and individuals (Arias, Eden, Fischer, Gorman, & Scharff, 2000; Bennis & Biederman, 1997; John-Steiner, 2000). Collaboration on line takes on many forms. It can be individual work within an interactive context such as the Math Forum’s Problems of the Week, in which an online mentor works with a student around his or her solution. Or, an individual can work with interactive site services, including the community of participants who populate the site, as in the case of the teacher working with the Math Forum. It can also be large or small group interactions that are formally structured and facilitated, as in the case of preservice teachers working within a STELLAR course. The Hawaii Networked Learning Communities, in turn, prioritizes deep engagement with a statewide school system to effect systemic improvement of science and mathematics education, using online collaboration technology to support the statewide community of educators engaged in this initiative. The emphasis to date has been on the pragmatics of this challenging application: gaining trust of the organizations and individuals within the over 30 schools statewide with which we work and developing the professional and leadership development model in conjunction with the software that supports it. Recently the network has reached the point where realistic evaluation can begin. Research foci for this project, including study of the expectations of new community members and what motivates their use or disuse of the online environment; the development of mentoring and collaborative relationships online and how these relationships effect change in practices within the organization; the use of an “artifact centered discussion” tool derived from previously funded NSF research, and use of a shared community database of educational resources.

Similarly, design support for collaboration takes many forms. It can be supported or hindered by a facilitator’s guidance, and by design of online tasks, tools and representational support systems (Suthers et al, 2004). For example, the Virtual Math Teams project in which teams of students are organized online to work on sets of problems (e.g., problems that prepare them to take SATs, or rich non-routine projects), required thoughtful design of site services to promote productive interactions with others. A major design consideration in this instance was what feedback was required to support learners to continue working in the face of frustrating situations.
Whether or not participants are reflectively aware of the community-based aspects of their work online, the emergence of overlapping knowledge-building communities is a critical development in the Web-based universe. Helping this universe develop through scientifically-grounded design is an important part of our research agenda. If we assume that world of working and living relies on -- interdisciplinary and cross-cultural collaboration, creativity, definition and framing of problems, dealing with uncertainty and change, and distributing cognition across people and tools -- then the online learning environments we create need to prepare and support learners to be productive in such a world (Fischer & Derry, 2005).

Cognitive, Affective and Social Components of Engagement

Central to reform efforts in both mathematics and science are the goals of enabling learners (students, teachers, etc.) to make connections to and generate strategies for working with the tasks that they are presented (Ginsberg, 1998; Kuhn, 1989; Schauble & Glaser, 1990; Schoenfeld, 1992; Strauss, 1998; Tweney, 2001) and concern for the context, or conditions, that are needed to provide such support (Crowley & Schunn, 2001; RAND Mathematics Study Panel, 2003). Dewey (1914) in his now classic work, *Interest and Effort*, points to the power of interest to support students to engage, or make connections, to materials to be learned. He says that a person can not be made to have interest, but can be supported to develop interest. He also observes that where there is interest, effort follows.

In a forthcoming review of the literature, Hidi and Renninger (2005) note that interest—the predisposition to reengage particular content over time—differs from other motivational variables in at least four ways. First, interest has both affective and cognitive components, a position supported by neuroscientific research (LeDoux, 2000a, 2000b; Panksepp, 1998, 2003). Second, both the cognitive and affective components of interest have biological roots (Hidi, 2003; Davidson, 2000; Panksepp, 1998). Third, interest is an outcome of interactions between a learner and particular content. Finally, interest is always content specific rather than applying across all activity.

Interest has been found to have a significant impact on learners’ attention (Hidi, 1990, 1995; Hidi, Renninger, & Krapp, 2004; McDaniel, Waddill, Finstad, & Bourg, 2000; Renninger & Wozniak, 1985; Schiefele, 1998), goal setting (Harackiewicz & Durik, 2003; Harackiewicz, Barron, Tauer,
Carter, & Elliot, 2000; Pintrich & Zusho, 2002; Sansone & Smith, 2000); and levels of learning (Alexander, 1997; Alexander & Murphy, 1998; Hoffmann, 2002; Koeller, Baumert, & Schnable, 2001; Krapp & Fink, 1992; Renninger, 1989, 1990; Renninger, Ewen, & Lasher, 2002; Renninger & Hidi, 2002; Sadoski, 2001; Schiefele, 1999; Schiefele & Krapp, 1996; Schraw & Dennison, 1994; Wade, Buxton, & Kelly, 1999).

Because interest exists in the interaction between learners, particular content, and the social context in which the learning occurs, interest is a variable that can be impacted by changes in support or feedback and features of learning contexts including opportunities to work with others (Renninger, Sansone, & Smith, 2004). Interest appears to develop through phases that begin with the triggering of interest and can lead to well-developed individual interest over time (Hidi & Renninger, 2005). Although not well understood, shifts in interest over time appear to be characterized by the changing relation between positive affect and opportunities to develop and/or deepen knowledge.

While students need to have positive feelings about an activity if they are to think that it is “cool” and worth trying (Resnick, Rusk, & Cooke, 1998); activities need to be both appealing and substantial if learners are to continue to work with and learn from them over time. Importantly, interest can be supported to develop and with support, it can deepen over time (Renninger, 2000). Given that the online environment is new and affords numerous and evolving opportunities for learning, it has attracted the attention and use of a wide-range of learners. The goal for these learners is that they will continue to engage over time, exerting effort to make connections to and generate strategies for working with content that may be more difficult to learn, or less accessible, in other contexts. Not only does their learning need to be promoted and sustained, but the staff members of communities of which they are a part need information about the ways in which they and their design can help. Because it gates attention, and with support can develop and/or deepen over time, the variable of interest appears particularly promising for exploring the affective, cognitive, and social components of engaged learning in online communities

Research Agenda

The online environment is new and it needs to be fully understood. It is being used widely and its potential for supporting individual learners to learn through collaboration in small groups and/or in the larger context of community that is associated with sites needs to be examined carefully. The
context of the Web affords possibilities for archiving and studying aspects of learning that have not before been available and subject content (e.g., mathematics) that has not been known to exist.

The complexity of interacting factors and the differences among the individual, small group, and community dimensions of online learning have not been systematically studied in terms of participants’ learning and how learning can be supported. Experiments to date have primarily been descriptive formative evaluation studies focused on particular contexts that are not generalizable. In Catalyst discussions and the proposed studies of a SLC, commonalities of findings across online contexts need to be identified, methods should draw on multiple disciplines, and systematic studies that allow comparison across forms of online learning, and include control groups need to be designed and conducted.

Thus, key considerations for the proposed Catalyst year include characterization of what presently is understood about the unique aspects of learning dynamics in online communities, including (a) hypotheses about the way in which learning interacts and evolves for individuals and for groups in the development of online communities and (b) information about what practitioners and developers need and want to understand. What is presently understood as well as the hypotheses to be developed about learning online will be framed around the center’s key research questions from the vision:

- When deep and engaged learning occurs in online communities, what is taking place and how can it be supported and sustained?
- What new forms of joint activity online are responsible for different aspects of human development, and what are the psychological and social consequences of these forms of activity?
- What are the necessary conditions for the development and sustainability of online learning communities?
- What is the relation between learner development and the online collectivities in which learners participate?

It is expected that research slated for study by a SLC will target (a) the cognitive, affective and social relations between learners and the communities in which they participate (individual, small group, and/or community) and (b) the forms of joint-activity that learners engage online, including the psychological and social consequences of these forms of activity. Sampling needs to be purposeful and independent variables need to include, at minimum, gender, minority status, and age of participant.
Specific research goals to be developed during the catalyst year that will inform the center are listed below:

- Develop a framework for engaged online learning based on insights from the learning sciences and innovative use of new media.
- Identify success models for engaged online learning in various STEM fields.
- Develop and study prototype online learning environments that:
  - Engage students in educational experiences that will help qualify them for and support them in successful STEM careers.
  - Scaffold students to learn through participation in technically, scientifically, socially, and artistically important inquiry and design.
  - Improve teaching and learning-environment design at all levels by fostering engaged individual and collaborative learning.
  - Provide models for broadening the sectors of the nation’s population that aspire to and participate in STEM education.
- Understand what it means to be connected and what it means to collaborate for students in school, where they will be using powerful mobile devices for learning, entertainment, socializing, etc.
- Creating a center where the scientific principles underlying this new form for interaction is studied in all its forms.

The Catalyst year then will involve the bringing together of the existing data and studies on online learning communities in order to consolidate the descriptive phase of research in this area. The research teams will then frame the different learning theory perspectives and appropriate methodologies for those perspective in order to define the work in this field and create and organization structure around the further development of that work. Finally these intellectual activities will be brought into dialogue with the discussion about the organization of the Center for Engaged Learning in Online Communities in order to frame the structure of the center and make sure it maps onto the intellectual imperatives in this field. This work will ultimately contribute to the development of the center proposal.
Project Plan & Timeline

The Catalyst Project has the central goal of collaboratively developing a research agenda for extending the frontiers of knowledge on engaged learning in online communities. There are three primary deliverables:

The creation of an American network of researchers committed to defining and carrying out this research agenda. This network will coalesce into an online learning community. It will be effectively connected to research centers and networks in other parts of the world. It will have its own identity and infrastructure.

A proposal for a Science of Learning Center for the study of engaged learning in online communities. This proposal will describe an appropriate intellectual, organizational, technical and physical infrastructure to support rigorous, scientific study of this topic. The Center will bring together researchers with diverse, multidisciplinary approaches, in partnership with active online communities and schools across the country and around the world.

A special issue of the *International Journal of Computer-Supported Collaborative Learning (ijCSCL)* on the topic of engaged learning in online communities. This will be an important public presentation of the research agenda, reviewing the state of the art and motivating the agenda. The articles in this issue will be scholarly and peer-reviewed. Drafts of the articles will be presented at relevant international conferences: CSCL, CSCW, CRIWG, AERA, EARLI.

These deliverables will be developed through an iterative process of meetings and online collaborations involving the PIs, lead researchers, collaborating researchers and international collaborators. The process will be designed to foster partnership-building and interdisciplinary investigation of central research issues.

Following is a chart summarizing this process during the 18 month project period:

<table>
<thead>
<tr>
<th>#</th>
<th>month</th>
<th>activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>June 2005</td>
<td>Preliminary informal meetings with IKIT and others at CSCL 05 in Taiwan. PI visit to NIE in Singapore.</td>
</tr>
<tr>
<td>1</td>
<td>July</td>
<td>Meeting of Catalyst PIs: define 5 workgroup focal areas.</td>
</tr>
<tr>
<td>2</td>
<td>Aug</td>
<td>Invite lead researchers to join workgroups, approx. 5 people in each of 5 online groups. PI meets with</td>
</tr>
<tr>
<td>Date</td>
<td>Event</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>3 Sept</td>
<td>Workgroups hold meetings to outline whitepapers. Website is established. PI keynote presentation at CRIWG conference in Brazil.</td>
<td></td>
</tr>
<tr>
<td>4 Oct</td>
<td>Each workgroup drafts a whitepaper on its focal area.</td>
<td></td>
</tr>
<tr>
<td>5 Nov</td>
<td>Major meeting held to outline Center research agenda based on whitepapers.</td>
<td></td>
</tr>
<tr>
<td>6 Dec</td>
<td>Draft Center research agenda. Select additional lead researchers for Center to complete disciplinary coverage.</td>
<td></td>
</tr>
<tr>
<td>7 Jan. 2006</td>
<td>Draft preliminary Center proposal. Select PIs for Center.</td>
<td></td>
</tr>
<tr>
<td>8 Feb</td>
<td>Submit preliminary Center proposal.</td>
<td></td>
</tr>
<tr>
<td>9 Mar</td>
<td>Hold workgroup meetings to revise Center proposal.</td>
<td></td>
</tr>
<tr>
<td>10 April</td>
<td>PIs of Center meet to critique and strategize proposal. Present research agenda at AERA 06 conference.</td>
<td></td>
</tr>
<tr>
<td>11 May</td>
<td>Draft final Center proposal.</td>
<td></td>
</tr>
<tr>
<td>12 June</td>
<td>Submit final Center proposal; present research agenda at International Conference of the Learning Sciences (ICLS 06) at Indiana.</td>
<td></td>
</tr>
<tr>
<td>13 July</td>
<td>Outline journal special issue papers.</td>
<td></td>
</tr>
<tr>
<td>14 Aug</td>
<td>Draft journal special issue papers.</td>
<td></td>
</tr>
<tr>
<td>15 Sept</td>
<td>Circulate, review and critique journal special issue papers.</td>
<td></td>
</tr>
<tr>
<td>16 Oct</td>
<td>Present research agenda at CSCW 06 conference.</td>
<td></td>
</tr>
<tr>
<td>17 Nov</td>
<td>Final drafts of journal special issue papers.</td>
<td></td>
</tr>
<tr>
<td>18 Dec</td>
<td>Prepare Center start-up. Submit journal special issue papers to ijCSCL.</td>
<td></td>
</tr>
</tbody>
</table>

The planning process will take place at the individual, small group and community levels simultaneously. Although ideas and documents will be
Proposals for Research
circulated widely though a wider research community, five small groups will take the lead in focusing the work and developing core working documents. The five Catalyst PIs will be individually responsible for organizing and facilitating these groups. Each workgroup will concentrate on a broad area, such as the effect of individual cognition and affect on participation, the knowledge-building function in small groups or the impact of diversity on learning in online communities.

Initially the PIs will continue the process of collecting key research questions from both participating researchers and from leading practitioners in the online communities of the partners. The first set of working groups will pull together critical findings, important unanswered questions, and perceived needs or opportunities for collaboration from other fields. These will form the basis of whitepaper drafts articulating the research agenda from each focus. The larger conference will bring together representatives from the working groups and other leaders in the field to react to the whitepapers, along with sample data to stimulate ideas for shared infrastructure for data collection and analysis. The conference will identify gaps in the research considered so far, critique priorities, and build connections across the working groups and to other relevant research, including the generation of proposals for possible studies to address the emerging issues. The PIs will draft a Center research agenda based on the conference and working group results to date. The next set of smaller working groups will be reconfigured to focus each on a core research program and the related coordinating functions of the proposed Center. The evolving PI team for the Center integrates working group output and takes a draft Center proposal back out to the field looking for theoretical and empirical studies that tie together Center foci. Strong paper concepts that emerge from the whitepaper and proposal process will be cultivated for development and submission to the journal after the proposal submission.

Web technology will be used for on-going communication within the Catalyst project. A project website will include support for email lists, threaded discussion forums, chat rooms, videoconferencing, wikis, a document repository and web pages. Each month, work on the project will be documented on the website, so that all participants can access and comment upon work product and draft deliverables. Infrastructure support for the Catalyst project will be supplied by the Math Forum and the project will be coordinated by the Math Forum Director and the PI.

The publication of a special issue of *ijCSCL* and presentations at relevant conferences will motivate researchers to think deeply about the research agenda and to review the state of the art in a thorough way. It will ensure
that the work of the Catalyst grant has a meaningful outcome and furthers progress on the topic of engaged learning in online communities regardless of eventual Center funding. *ijCSCL* is an appropriate venue for this topic (see [http://ijCSCL.org](http://ijCSCL.org)), and a number of Editorial Board members are involved in the Catalyst project, including the PI, who is one of the Executive Editors along with one of the Catalyst international collaborators. *ijCSCL* is an international journal and includes all the contacts for the international networks associated with this project (see below) on its Editorial Board.

The Catalyst project brings together an exciting mix of researchers (see below). The project is designed to mold this collection of people and small research centers into an integrated community — itself an online learning community. The nature of the problem being addressed by this Catalyst project requires a deeply interdisciplinary and tightly collaborative process. The PIs are all experienced in coordinating interdisciplinary efforts. They will use the large and small meetings as well as the online communications to intertwine and merge different perspectives to arrive at cross-perspective issues and views. The emergent research agenda must be more than the sum of participants’ individual professional agendas — and the Catalyst project will be structured to achieve this through its community-building emphasis.

**Project Personnel**

**Principal Investigators**

The PIs are learning sciences researchers associated with four active online learning communities. They have each been involved with a variety of online learning community projects. Among them, they have studied a diversity of different kinds of efforts (see Background and individual bios).

- Gerry Stahl, Drexel University, Virtual Math Teams Project at the Math Forum @ Drexel
- Sharon J. Derry, University of Wisconsin, Wisconsin Center for Education Research
- Mary Marlino, UCAR, Digital Library for Earth System Education Program Center
- K. Ann Renninger, Swarthmore College, Math Forum @ Drexel
- Daniel D. Suthers, University of Hawaii, Hawaii Networked Learning Communities
The PI team will be supported by Stephen Weimar, Director of the Math Forum @ Drexel, who brings a deep experience with K-12 education and organizational development.

**Lead Researchers**

These are the researchers who will most likely be involved on the focal workgroups, drafting the whitepapers and journal articles. Most have confirmed their involvement (see letters in Appendix), although the list is still somewhat in formation. Many of these people are the pioneers of the field of online learning. Some continue to be leaders, while others are conducting rigorous research on related topics that are central to issues of engaged online community and its development. They bring diverse methodologies from a spectrum of fields. Although most may consider themselves interdisciplinary, they were trained in education, psychology, computer science, information science, anthropology, philosophy, social sciences, cognitive sciences, learning sciences.

Robert B. Allen, Drexel University  
Sasha Barab, Indiana University  
Amy Bruckman, Georgia Tech University  
Paul Cobb, Vanderbilt University  
Kevin Crowley, University of Pittsburgh  
Danny Edelson, Northwestern University  
Clarence (Skip) Ellis, Univ. of Colorado  
Gerhard Fischer, University of Colorado  
Mary Gauvain, Univ. California at Riverside  
Geri Gay, Cornell University  
Ricki Goldman, New Jersey Inst. of Techn.  
Rogers Hall, Vanderbilt University  
Judy Harackiewicz, Univ. of Wisconsin  
Starr Roxanne Hiltz, New Jersey Inst. Of Technology  
Cindy Hmelo-Silver, Rutgers University  
Jim Kaput, Univ. Massachusetts, Dartmouth  
Mick Khoo, University of Colorado  
Wesley Shumar, Drexel University  
Elliot Soloway, University of Michigan  
Nancy Songer, University of Michigan  
Rand Spiro, Michigan State University  
Tamara Sumner, University of Colorado  
Stephen Weimar, Math Forum @ Drexel  
Margaret Wilsman, University of Wisconsin
Collaborative Researchers, American

A list of about two dozen other researchers will also be involved in this Catalyst project. They may attend the major project meeting, participate in online discussions, contribute to the whitepapers and co-author journal articles. Most of the people listed above work in research centers with associates, students and networks of colleagues. So this Catalyst will engage a critical mass of the relevant research communities.

International Collaboration

The Catalyst project on engaged learning in online communities – and its successor SLC Center – will be the US partner in an international collaboration. There are several reasons for such a collaboration:

- Online communities can easily extend across national boundaries
- Collaborative learning is enhanced by the inclusion of international perspectives
- A number of countries are more advanced than the US in the fostering of collaborative learning
- A number of leading learning scientists interested in online communities are located abroad
- The research communities of the learning sciences and computer-supported collaborative learning (CSCL) are international in character
- International collaboration will enhance the dissemination and impact of the Center’s findings

The PIs have well-established international ties in the relevant international research community. A research workshop associated with the PI’s VMT project at the Math Forum last summer brought together 36 researchers from 10 countries, including members of some of the networks collaborating on this project. The project will formally collaborate with the following research networks:

- The Kaleidoscope Network of Excellence in the European Union, a network of learning researchers in Europe that includes a 300-member CSCL SIG. Contact: Barbara Wasson, University of Bergen, Chair of the CSCL SIG.
• The Knowledge Management Research Center, a major social science research center conducting basic research on learning. Contact: Friedrich Hesse, University of Tübingen, Director of KMRC in Tübingen, Germany.

• CRIWG, a network of Latin American CSCL and CSCW researchers. Contact: Hugo Fuks, Catholic University of Rio de Janeiro, Brazil.

• Learning Sciences and Technologies Group in the National Institute of Education (NIE). Contact: Chee-Kit Looi, Nanyang Technological University, Singapore.

• Institute for Knowledge Innovation & Technology (IKIT), an international network growing out of the work on CSILE. Contact: Marlene Scardamalia, University of Toronto, Canada.

Diversity

One of the co-PIs (Suthers) conducts research on online learning in the context of his own teaching at the University of Hawai`i, a minority-serving post-secondary institution located in an EPSCoR state. He is also engaged in studying learning in the context of online communities of K-12 students and teachers within the “Hawai`i Networked Learning Communities” (HNLC) project, a partnership between the Hawai`i Department of Education and the University of Hawai`i at Manoa (NSF Rural Systemic Initiative, Cooperative Agreement #0100393). The overall goal of the HNLC is to improve science, mathematics and technology learning in the K-12 rural schools. The project focuses on professional development and leadership development for teachers and administrators in the state’s rural and remote schools, most of which are located on the outer islands of the state. Of the participating schools, 14 schools are located on the island of Hawai`i, with eight schools on Kaua`i, six on Maui, three on O`ahu, three on Moloka`i and one on Lana`i. To reduce the isolation of the teachers and administrators in these remote and rural schools which are even further isolated because of the island configuration of the state, the project utilizes internet technology in the form of a “Virtual Community Center” (hnlc.org) to support collaboration between a distributed statewide community of teachers and students.

The student population is culturally diverse, and there is no single “majority” ethnic group among students. The largest populations are Filipino ranging (across three cohorts) between 27-29% and the Part-
Hawaiian student group (23-25%). These two groups combined with the Native Hawaiian students (4-6%) represent just under two-thirds of the student population and these students are underrepresented in the target courses in math and science. Furthermore, HNLC reaches an under-served population as defined by socioeconomic status. Currently, 6.2% of the students are Limited English Proficient and 48.1% eligible for free or reduced price lunch. In addition to the issues of size and rural isolation, the September 2003 release of the state data on No Child Left Behind (NCLB) Annual Yearly Progress (AYP) status revealed that 85% of project schools did not meet AYP compared to a state-wide rate of 66%. More significantly, nine of the 20 schools or 45% are in “corrective action” or “planning for restructuring” compared to only 24.5% of state’s schools. In combination with low student performance and small size, many HNLC schools are impacted by low teacher certification rates and poor economics. These findings suggest that the project’s rural schools are among the most impacted in the state. As a result, the HNLC project provides a challenging testbed for the role of online communities in helping schools meet performance objectives.

Evaluation and Assessment

In preparing a Center proposal it is essential to assess the effective working relations of the collaborators and the level of productivity and quality in the results. Sara Kiesler's studies of the KDI program confirm the importance of planning and regular onsite meetings for geographically spread collaborations. Derry, Weimar, Marlino and others involved have significant experience studying effective collaborations and in organizational development and strategy. This team will conduct a survey of project participants in December and then again in October of the following year concerning levels of participation, the quality of the meetings, the quality of the work between meetings, the quality of the results and the emerging design and focus of the Center. An internal review will also be conducted to analyze project activities for the level of interaction, follow through and continuity, adherence to the timeline and plan, and the contributions of various means of collaboration to the overall effort. The team will be able to draw on participating ethnographers such as Wes Shumar and Mick Khoo who have experience studying the development of project groups. These results will form the basis for the design of collaboration functions supporting the Center.
References


Harackiewicz, J. M., & Durik, A. M. (2003, August). Task value in the college classroom: Predicting goals, interest, and performance. In M. Niemivirta (Chair), *Advances in achievement goal research: The role of moderators and mediators*. Symposium conducted at the 10th Biannual Meeting of the European Association for Learning and Instruction, Padova, Italy.


Hidi, S. (2003, August). *Interest: A motivational variable with a difference*. Plenary address presented at the 10th Biennial Meeting of the European Association for Learning and Instruction, Padova, Italy.


Proposals for Research

learning: A project of the National Council of Teachers of Mathematics (pp. 334-370). New York: Macmillan.


ITR: Catalyzing & Nurturing Online Workgroups to Power Virtual Learning Communities

Project Summary

The Math Forum website (www.mathforum.org) combines educational activities, a digital library of resources, automated and manual mentoring, discussion forums and other mathematics-related services for students, teachers and the public. As an intellectual home to about a million people, it demonstrates how the Internet can host large global learning communities. However, learning here has been primarily oriented toward individual learning. The Math Forum now aims to bring together some of its visitors with similar interests to work on common issues: students exploring a math issue together, teachers developing curriculum, or technologists and educators designing interactive math applets and support tools.

Collaborative learning in small workgroups can be particularly effective in motivating interest in math and in building and communicating deep understanding. A proliferation of small groups will heighten the sense of a vital community and increase its ability to become self-sustaining and vigorous. The groups will help people increase their community participation and their interest in mathematics.

Problem: How can one catalyze the formation of online workgroups? If there are a couple thousand users logged into the Math Forum when you log in, how can you automatically be put in touch with an optimal selection of 4 to 6 of those people whose interests and abilities best complement yours? Once formed, how can your group be nurtured with online tools, processes, structures and mentoring to maximize group success and collaborative learning? How can networks of different kinds of small groups support one another within the context of the larger community?

The Project will investigate these questions and related issues through a series of pilot studies, controlled experiments, prototypes and field studies using group-formation and group-scaffolding software that is designed, implemented and assessed in collaboration with an international,
multidisciplinary group of leading HCI, CSCW and CSCL researchers. In particular, three different kinds of groups will be formed and supported: (a) groups of students who visit the site and work on a “collaborative problem of the week”, (b) groups of teachers, student teachers and mentors who develop new problems, and (c) multidisciplinary groups of international researchers and developers who design and assess technologies and interventions.

**Intellectual Merit:** This Project explores a primary open challenge of the Internet – with detailed and rigorous methods, under controlled and real-world global conditions: how to foster effective collaborative online learning. It joins the multidisciplinary expertise of the international CSCL community with the practical success of the Math Forum to study how to mediate the growth of a large virtual learning community, and to design, develop and assess tools for the automated support of small workgroups acquiring, managing and negotiating knowledge.

**Innovation in IT:** An unfulfilled promise of the Internet is to bring together people who do not know each other or live close by, but who could benefit from interacting within knowledge-rich contexts. This Project addresses core issues of computer support for collaborative learning (CSCL): how best to form and structure intimate learning workgroups within global knowledge-building communities and how to effectively scaffold their interactions.

**Integration of Research & Education:** The Math Forum is a major practical success of prior NSF research, forming a virtual community of about a million students, teachers and mathematicians. This Project will systematically initiate and support efforts to form small collaborations within the large body of users who now interact as individuals with the site. This fundamental research into innovative support for small group collaborative online learning will take place within a vibrant and realistic large-scale context and will impact all levels: student motivation and learning, teacher development, and community evolution – generating a new model of global virtual learning communities, incorporating the power and motivation of small-group collaboration.

**Broader Impacts:** The Math Forum model, with automated formation of small groups and support for interactions developing deep understanding of mathematics, will be suggestive for virtual learning communities in other domains, taking advantage of other digital libraries. This model provides opportunities for students and teachers excluded from collaborative learning due to geographic isolation, disadvantaged schools, physical disability, discrimination and other physical or social factors. The model stimulates
both student motivation and teacher development, transforming interest in mathematics from a social stigma into a bridge to global friendships.

**Integrating Diversity:** A central Project hypothesis is that groups integrating diversity of all kinds learn better.

**International Collaboration:** The Project builds on the PI’s prior work on an EU grant. Core aspects of the Project – including technology design, pedagogy and assessment – will be conducted by workgroups of American and European leaders of the CSCL community in collaboration with Project staff. Annual workshops at international conferences will bring these collaborators together with each other and with wider international audiences.

### Project Description

#### 1. Vision

It is January 2009 and the proposed project has just ended. Tanja is home schooled in up-state New York because of a physical disability; Sarah lives on a remote Navaho reservation; Damir attends school in Croatia. They each read the same “Problem of the Week” (POW) on the Math Forum (MF) website and became interested in it. The problem is to specify a general equation or algorithm for saying how many squares a straight line segment on graph paper will go through, given the coordinates of the line. Based on previous visits of Tanja, Sarah and Damir individually, the new MF website software determines that they have a mix of interests and skill levels that might work well in a small group solving this problem together. MF invites them and a couple more students to work together. Tanja, Sarah and Damir respond and find themselves together in MFCE, the online Math Forum Collaboration Environment. MFCE helps them to coalesce into an on-going group to work on this problem; to communicate both synchronously and asynchronously; to represent the problem and its features; to store, reflect upon and reorganize their collaborative ideas; to negotiate a group response to the problem; to document how they arrived at their response; to submit their response to MF; to receive immediate feedback; and to decide if they want to continue collaborating.

Sandra, a teacher who has used MF with her math classes for years, reads the solution submitted by Tanja, Sarah and Damir and starts to think about a related problem, which she posts to a MF discussion. A number of other
teachers respond with interest, and the MF website software invites them to work together with a MF staff person to develop this idea into a publishable POW. They use MFCE to collaborate, reviewing several of the responses to the previous POW and eventually releasing a new problem that asks how many regions are formed by connecting N points on a circle to each other.

MF staff notice that a number of recent POWs involve drawing simple line figures and counting features such as vertices and regions. They request MFCE to set up a work group with several of their technical collaborators around the world who might be available and appropriate. Together, the people who respond form a group, specify user requirements, brainstorm designs, develop prototypes, conduct user testing and develop a new tool for MFCE. This tool allows student groups to collaboratively sketch representations that help them visualize and communicate about features of 2D drawings.

Back to the present. This Project will explore the potential of the Internet to bring together small groups of people like the three groups described above, who share interests and skills that might allow them to learn collaboratively and to build knowledge together. By the start of the grant period, MF is projected to be serving over a million students per month. This means that by the end of the grant period it will be common to have over a thousand math-oriented people online with MF at the same time during peak usage. This is a rich pool for forming compatible small groups at various levels of mathematical interest for collaborative learning.

It is clear from current learning theory that collaborative learning is an effective way for many people to learn, particularly people who tend to be left behind in classroom situations (Johnson & Johnson, 1989). It is also clear from experience that an approach that emphasizes discourse and interpersonal interaction helps to build deep understanding of mathematical principles – rather than simply exercising rote memorization – in keeping with contemporary pedagogical priorities (NCTM, 2000; Renninger & Shumar, 2002). What is not clear is whether software can be developed to match people who are using the Internet based on their known interests and skills, or whether these people can be formed into effective groups for collaborative learning. There has been almost no relevant research about online group formation (Haake et al., 2003; Wessner et al., 2002; Wessner & Pfister, 2001). Even in face-to-face settings, there are many open questions about how to form effective learning groups and then how to structure their collaborative tasks (Stahl, 2000b). Finally, only rather primitive software is available to support collaborative knowledge building (Stahl, 2002d). This project will explore the theoretical, technological and
pedagogical issues and will systematically design, implement and assess an integrated approach to foster the building of mathematical knowledge in virtual groups.

2. Innovation & Significance

The long-heralded promise of the Internet was that people could not only access the whole world of information from any location, but also that they could meet people who shared their interests and could explore ideas together (Bush, 1945; Engelbart, 1995; Hiltz & Turoff, 1978; Rheingold, 1993). With the exponential growth and the specialization of the world’s knowledge in mathematics and science, for instance, it is unlikely that someone with a particular momentary interest would happen to know many other people in their physical neighborhood with that interest at that time. Yet, there are currently few examples of technologies that fulfill this promise of the Internet and allow one to find and work with some of the many people who might share one’s interest across the nation or around the world. This project proposes to develop such a technology and to demonstrate its effectiveness in the learning of school mathematics.

Information technology (IT) to date has transformed how individuals work and learn. The Internet has leveraged the productivity of desktop support by allowing individuals to communicate ideas (email, messaging, chat, newsgroups, video-conferencing) and to share information (websites, digital libraries, shared repositories). But there has been little progress toward supporting the intense interactions of spontaneous small group collaboration that builds shred knowledge. The motivation for this Project is that support for small group collaboration may yield the next major benefits of IT for working and learning. Progress in this direction may consist largely of adapting and packaging technologies that are now within reach – for the tough research and implementation issues are more social than purely technical. Careful, detailed, rigorous study is needed of particular technologies in specific social contexts. This Project will study alternative IT solution to catalyzing and nurturing several kinds of workgroups within the context of the MF virtual community.

The Math Forum (MF) is an NSF-supported organization and mathematics website that offers a variety of services, primarily to students and teachers interested in topics of mathematics commonly encountered from kindergarten through calculus courses. It currently receives over 800,000 online visitors a month. Most of these are people who are solving mathematics problems on an individual basis. MF is interested in supporting
more collaborative approaches to learning – not only for student visitors working on the popular Problems of the Week (POWs), but also for teachers developing curriculum and for people associated with MF who are developing new resources such as new POWs. MF would like to harness, extend and apply IT to foster and support the formation of small groups of people to explore topics in mathematics.

Computer support for collaborative learning (CSCL) is an established research field, offering technology, theory and pedagogy. Although there are comprehensive CSCL systems to support classroom learning (e.g., WebCT, Blackboard, Lotus LearningSpace, Knowledge Forum / CSILE, WISE / KEY, Synergeia / BSCL), these systems do not include support for the formation of small workgroups based on criteria of compatibility – even within classrooms, let alone in larger, more amorphous communities. The proposed project will support the formation and subsequent collaborative learning processes of a variety of specific types of small groups drawn from the MF community.

This project will approach group formation in a systematic way, developing theory, technology and pedagogy that are integrated together. The collaborative learning theory will describe the phases of group interaction, such as: group formation, task specification, brainstorming, proposals, negotiation and publication of results. These reflections will build not only on theoretical frameworks common in the CSCL literature, but specifically on past studies of the MF virtual community (e.g., Renninger & Shumar, 2002; Renninger et al., 1989). The computer support technology will provide a comprehensive environment within which virtual groups can successfully pass through these phases. It will be developed using best practices of user-centered human-computer interaction and extensive iterations of user testing within MF. The pedagogy will describe the nature of appropriate group membership criteria, problem characteristics and process facilitation. It will involve a reorganization of the usual MF process oriented to individual learning into one oriented to collaborative group learning, as described in the following paragraph.

The MF community can be described as a pyramid, with a broad base of individual first-time visitors, followed by successive layers of: loyal readers, contributors, facilitators and finally a small staff. While MF began in 1996 as a top-down, funded effort, its goal is to create a self-sustaining community where ideas, POWs, and activities flow up from the bottom. The formation of virtual groups will be a major means for achieving this, by encouraging and supporting people to move up the pyramid from occasional visitor to co-designer of MF services. In particular, groups will be formed at
three levels in this project: (a) visitors/readers who want to solve collaborative POWs (cPOWs), (b) contributors/facilitators who formulate new cPOWs and (c) facilitators/staff who maintain and extend the MF technical infrastructure. (a) New problems will be offered that are suited to collaborative learning, stressing richer, more open-ended topics designed to foster discussion of deeper mathematical understanding. (b) These new cPOWs will be designed by groups formed of contributors, who will have access to an extensive digital library of past MF problems, math-related applets and studies of responses to previous problems. (c) Groups of facilitators, programmers and MF staff will form to maintain MF’s mathDL and related services, including the production of new mathlets (programmable applets for computers and handhelds), in response to the needs of groups (a) and (b).

The project is designed to be highly iterative, so that the different aspects of the project can evolve in response to each other. The project starts on the basis of considerable experience with POWs solved by individuals – generally working within supportive communities of classrooms or MF mentoring, but without collaborative learning groups. The first year focuses on micro-analysis of this experience, incorporating previous studies of the MF virtual community, but clarifying the theoretical and pedagogical issues as well as the specific user requirements for the technology. In the second year, technological supports are gradually introduced to support group formation and knowledge building. An initial set of cPOWs that have been meanwhile adapted from old problems is used. In the third and fourth years, groups at all three levels are active; well-defined experiments and formative evaluations drive revision of the technologies and practices. In the final year, the project will observe the stable functioning of self-forming groups at all three levels and will evaluate the success of people finding compatible group partners and learning mathematics collaboratively. Logging of group interactions as well as MF hits will allow for careful quantitative and qualitative evaluation of the theory, technology and pedagogy.

3. Theory

Collaborative learning, as understood in this proposal, involves a focus on the group level of analysis. Of course, construction of knowledge by a group can also be seen as co-construction by the individuals in the group, and the building of knowledge by the group has direct implications for learning by the participants. However, it is also true that the group can produce knowledge that none of its members would have produced by themselves.
and it is true that, for instance, the meaning of things said in group discourse is defined by the group interaction itself rather than simply by ideas in the minds of individuals (Stahl, 2003c). Although this project will also be concerned with the learning of individual participants, its focus will be on the building of knowledge by small groups. This emphasis is consonant with theories from cognitive science, communication theory, anthropology, education and CSCL, such as situated action (Suchman, 1987), activity theory (Engeström, 1999), situated learning (Lave & Wenger, 1991), distributed cognition (Hutchins, 1996), etc.

Traditional learning theory assumes that learning happens entirely in the mind of the individual, it can be led or facilitated by a teacher, the content is to some extent irrelevant (in that one can learn anything) and that this is a primarily cognitive act. But advances in learning theory have led to very different assumptions about the learning process. First is the realization that learning is a constant and ongoing process. People are always learning whether it is part of some specific curriculum or not. As Dewey pointed out the purpose of teaching and education is not just to help students learn but to create opportunities for experiences that lead to productive forms of learning (Dewey, 1938/1991). That means connecting new experiences with a persons prior experience base and creating opportunities for educative new experiences. In other words, creating a social context where new experiences can lead to the moral, emotional and intellectual development of the person. From Dewey to Vygotsky (Vygotsky, 1930/1978) one strand of learning theory has focused on the individual in social context and how that context creates opportunities for learning. While there have been tremendous advancements here in our understanding of learning, it is still often thought of as primarily something that happens to the individual (even if in a social context) and something that is primarily cognitive.

Very recent work on situated learning (Lave, 1991; Lave, 1996; Lave & Wenger, 1991) activity theory (Engeström et al., 1999) and cultural theories of learning (Cole, 1996; Holland et al., 2000) have moved thinking about learning from the individual to the intersubjective experience and from the cognitive to the whole person including affective dimensions of learning. This work has led to several important implications for understanding learning. First and perhaps most importantly, learning is a social process that involves individual interest, membership in a community where others are learning and engaged in productive social practice and that knowledge is built intersubjectively and shared among members of the community. Finally the constraints on learning and knowledge production are constraints that exist within the social system, its form of organization and patterns of interaction and not within the individual. These important theoretical
realizations about learning then have tremendous implication for collaborative learning and CSCL.

The emphasis on the group unit has methodological implications. The analysis of what takes place in project experiments will rely heavily upon interaction analysis (Duranti, 1998; Garfinkel, 1967; Heritage, 1984; Jordan & Henderson, 1995; Sacks, 1992) (Stahl, 2002) and community ethnography (Renninger & Shumar, 2002). These analyses will study in quantitative and qualitative terms how small groups function to mediate the building of a larger knowledge-building community and how groups engage in sequences of different kinds of interactions to build their knowledge. Tentative theories about how this knowledge-building process takes place will be subjected to empirical study, feeding back into revised formulations of a theory of collaborative knowledge building.

While we know a great deal more about the social nature of the learning process, many of our educational institutions continue to be structured in ways that assume learning is individual and a matter of transferring information from teacher to student. That sad fact means that many of the things students learn most effectively in school are patterns of resistance and lessons from the marketplace that appeal more to a social and collaborative form of learning. While some schools have begun to implement collaborative forms of learning there are real limitations due to the political context of local schools and the difficulty of organizing resources and groups that cluster together concentrations of expertise and appeal to individual’s interests. The Internet and digital libraries such as the Math Forum create an important strategic opportunity to bring a more collaborative community of practice to individuals who may be distributed geographically in different school and institutional sites. Therefore a critical next step in theoretical development is to figure out how to make online groups self-forming and self-replicating.

### 4. Pedagogy

The building of deepened understanding and increased knowledge of mathematics takes place in motivational community contexts, such as classrooms and research fields (Lave, 1991; Lave, 1996). Interactions within small groups can mediate effectively between individuals and these larger communities, providing supportive settings and engaging activities (Wenger, 1998). Small groups can build knowledge (collaborative learning) that draws upon and may extend the community knowledge while making it available to the individual participants who contribute to the group.
knowledge. According to theories of situated learning (Lave & Wenger, 1991), changing patterns of participation in which individuals become progressively more involved are important features of community learning; we have already seen signs of this taking place in the Math Forum virtual community in the documented example of Sonia and her son (Renninger & Shumar, 2002, p. 66 ff). This project will investigate the effects of online collaborative math learning by extending the services of the Math Forum and its growing community. It will explore the effect this has in drawing average or poorly motivated students into intellectual engagement, as well as involving students and teachers already excited about math in a larger community.

Mathematics is often thought of as the discipline of “the right answer.” A small group of teachers and Math Forum (MF) staff became uncomfortable with this designation because it can interfere with efforts to help students express their mathematical thinking, learn from mistakes, experiment effectively, and pursue their mathematical interests. They asked, how can we transform the student's question “Am I right?” into “How can I develop confidence and judgment that I am on the right track when working on a problem?” and “How can I know that I am improving my mathematical problem-solving and communication skills?” They decided that engaging students in discourse about mathematics was the way to go.

Discourse can make thinking public and create an opportunity for the negotiation of meaning and agreement (Bauersfeld, 1995). At the same time, discourse provides collective support for developing one's thinking, drawing it out through the interest, questions, probing, and ideas of others (Cobb, 1995; Krummheuer, 1995; Wood, 1995; Yackel, 1995), and discourse enables students to connect their own everyday language with the specialized language of mathematics (Vygotsky, 1934/1986). Articulating what they know allows students to clarify their own understandings. Through discourse, a teacher can better grasp the mathematical needs of the class: what the students know, misconceptions they may have, and how these might have developed (Resnick, 1988). Teachers and students gain perspective on their own thoughts through the attempt to understand the thinking of others, in the process laying the foundation for a supportive learning community (Brown & Campione, 1994).

Within the mathematics education community there is strong interest in the use of discourse for teaching and learning mathematics (Atkins, 1999; NCTM, 2000; Schifter, 1996). The teacher's role is described in broad terms as facilitative, to include listening carefully to students, framing appropriate questions, and mediating competing perspectives. Students are expected to
develop problem-solving skills: defining problems, formulating conjectures, and discussing the validity of solutions. Stigler and Hiebert (1998) report similar roles for teachers and students in mathematics classrooms in Japan, where mathematical discourse is an integral part of instruction.

The best way to foster domain-oriented discourse is to catalyze active small workgroups. In heterogeneous small groups, students are challenged to stretch and learn within their “zone of proximal development” “in collaboration with more capable peers” (Vygotsky, 1930/1978, p. 86). At the same time, the mentoring experience is productive for the “more capable” peer’s learning by teaching – and these roles are likely to reverse in other situations when the group members have complementary strengths.

5. Technology

Consider the Internet. It is a huge computational machine. It processes information reflective of the interests of millions of people. It is not simply a poorly organized repository of textual facts; it is the infrastructure of a global community. No one can navigate around it easily to find the particular things of interest to them and no one can relate to the world’s population in a human way. Yet, for most people, there must be gems of information hidden out there and potential friends or colleagues who could help them to make sense of those gems. The taming of the Internet’s informational and human vastness poses the technical challenge of our time. While much research is conducted on searching and organizing the information, surprisingly little has been done on bringing together groups of people on a human scale to use the Internet collaboratively.

Some organizations have explored systems for locating expertise within their staffs (Ackerman & McDonald, 1996); but the techniques for that do not transfer to the problem of finding people on the Internet with matching interests. There have been some experiments with social awareness, to display other people who are viewing the same web page at the same time (Graether & Prinz, 2001), but this hint is not enough to support group formation. A group formation project in Japan matched learning theories (Inaba et al., 2000; Supnithi et al., 1999), but not people. A prototype for group formation in Germany allowed students who knew each other to self-select groups (Wessner et al., 2002; Wessner & Pfister, 2001), but this approach does not scale to large groups who do not know each other personally. A spin-off of the German research is being expanded and developed for distance education, and the proposed Project will collaborate with that one (see section on International Collaboration).
The PI began exploring support for group formation while teaching an online HCI (Human-Computer Interaction) course for graduate students at Drexel. His students studied the issue and came up with several low-fidelity prototypes that they subjected to user testing. The PI developed an automated grouping agent, which he used to form work groups in a subsequent course. In both the student prototypes and the grouping agent, groups were formed based on specific criteria about the participants: their schedules, their interests and their skill levels. These pilot studies for the proposed project suggest the kinds of balance that should be sought in forming distributed groups. For instance, if synchronous communication is to be possible within the group – especially given different global time zones – members must have similar schedules. On the other hand, collaborative teams often work best when there is a diversity of perspectives and skills, along with a commonality of interests. Thus, a matching algorithm must optimize certain similarities and other differences. Diverse theories of collaboration stress the power of heterogeneity: cognitive dissonance (Festinger, 1957), perspectives (Boland & Tenkasi, 1995; Goldnam-Segall, 1998; Stahl & Herrmann, 1999), interdependence (Johnson & Johnson, 1989), zone of proximal development (Vygotsky, 1930/1978), cognitive flexibility (Feltovich et al., 1996). This Project will systematically explore the hypothesis that balanced heterogeneous small groups collaborate more effectively and will develop algorithms and prototypes to implement support for this.

The pilot study of group formation was conducted with a class using two different online collaboration environments: Blackboard and BSCL. Blackboard is a commercial system to support collaboration. It is used widely in university courses, particularly in the US. Blackboard can be extended (in Java) by third party developers using the Blackboard Building Blocks SDK (see http://buildingblocks.blackboard.com/bin/bbdn_info.pl). BSCL (Basic System for Collaborative Learning) is a system with collaboration support for classrooms that is similar to Blackboard (Stahl, 2002d). It was designed and developed by the PI and others in 2001/2002 as part of a European Union research project. BSCL is an extension (developed in Python) to BSCW (Appelt & Klöckner, 1999), a shared repository CSCW system widely used in European research and learning organizations. It is available for free to academic organizations. The PI has a license to develop it during the period of this Project (see Letter of Support in Supplementary Documentation).
The Math Forum has custom software (developed in a Perl-based environment) to support the virtual community and digital library of math resources and activities.

This Project will design, develop and evaluate software extensions to Blackboard, BSCL and Math Forum. This software will implement alternative approaches to group formation, discussion, shared representations, social awareness and knowledge management within the context of catalyzing and nurturing small groups within the Math Forum community.

6. Project Team

Information Science & Technology

Drexel University has a long history of technology leadership, dating back to the 1980’s when it was the first university to require entering undergraduates to have a PC.

Drexel University’s College of Information Science and Technology is rated the #1 graduate school of library science information systems by US News and World Report (http://www.usnews.com/usnews/edu/grad/rankings/lib/brief/infsp3_brief.php). This interdisciplinary College offers online and campus-based undergraduate and graduate programs in computer science (e.g., HCI, databases, software engineering) and library science (including digital libraries).

The PI is an Associate Professor in the College of Information Science and Technology. He brings a multidisciplinary background to the Project, with PhD dissertations in philosophy/social theory and computer science/AI (Stahl, 1975; 1993a). He has developed a series of collaboration support systems: Hermes (Stahl, 1993b), WebNet (Stahl, 2000a), WebGuide (Stahl, 1999a; 1999b; 2001; Stahl & Herrmann, 1999), BSCL (Stahl, 2002d; 2003b), and other educational software: Teachers Curriculum Assistant (Stahl et al., 1995a; Stahl et al., 1995b) and State-the-Essence (Kintsch et al., 2000; Stahl & dePaula, 2001).

The PI specializes in CSCL research, having published on CSCL theory (Stahl, 1993b; 1998; 2000c; 2002c; 2003a; 2003c) and the use of discourse analysis as an assessment methodology (Stahl, 2002a; 2002e; 2002f; Stahl & Sanusi, 2001). He was Program Chair of CSCL 2002 and Editor of the CSCL 2002 Proceedings (Stahl, 2002b). He is Workshop Chair of CSCL
2003 and Communications Chair and founding Board member of the International Society for the Learning Sciences (ISLS) (http://www.isls.org).

**The Math Forum**

The Math Forum was founded in 1992 as the Geometry Forum at Swarthmore College, expanded to The Math Forum in 1996, and funded in its development by the National Science Foundation. It has become one of the most successful applications of the Internet to education through the development of interactive services that bridge the higher education, K12, and industry communities. These services form the basis for a knowledge building environment that generates high quality mathematical content, supports student learning, integrates the benefits of technology with education, and is used for teacher professional development and pre-service teacher education. The Math Forum now comprises over 1.2 million pages of content, has over 2 million visits a month, receives up to 9,000 queries a month at its Ask Dr. Math expert service, and mentored over 27,000 students during the 2000-2001 school year through its Problem of the Week services. Among its current projects are two NSF grants, one focused on the use of online student mentoring programs in pre-service teacher education courses, and the other on the development of MathTools, a digital library for software in mathematics education, arithmetic-calculus.

**Education & Ethnography**

Drexel University also has a School of Education and a Department of Culture & Communication, both of which are represented in this Project. Prof. Wesley Shumar is a cultural anthropologist in the Department of Culture & Communication who specializes in educational anthropology and has conducted ethnographic studies of the Math Forum for many years.

**National and International Collaborators**

A unique feature of this Project is the involvement of leading national and international researchers. They bring expertise from a variety of relevant specialties and perspectives. Their participation will provide a natural means for sharing practical knowledge from Europe and the US as well as for disseminating the results of this Project across the nation and globe. To ensure a strong cadre of collaborators, the following researchers have already expressed strong interest in participating in the Project; others can join in the future:
**Americans:** Geri Gay (Cornell), Ricki Goldman-Segall (NJIT), Cindy Hmelo-Silver (Rutgers), Christopher Hoadley (Penn State), Timothy Koschmann (Southern Illinois U), Bonnie Nardi (Agilent), Leysia Palen (Colorado), Linda Puliam (California State U.), Mark Schlager (SRI).

**International:** Wolfgang Appelt (Fraunhofer-FIT, Germany), Hugo Fuks (Rio, Brazil), Joerg Haake (Distance U, Germany), Kai Hakkarainen (Helsinki, Finland), Thomas Herrmann (Dortmund, Germany), Jim Hewitt (Toronto, Canada), Victor Kaptelinin (Umea, Sweden), Anders Morch (Oslo, Norway), Wolfgang Prinz (Aachen, Germany), Volker Wulf (Siegen, Germany).

These individuals are established leaders in the HCI, CSCW and CSCL research communities, having made important contributions in theory, system design and assessment methodology. They all recognize the importance of collaboration, both in theory and in practice.

### 7. Prior Work

**The Math Forum**

*REC-9618223, $971,300, March 1999 to February 29, 2000*

The Math Forum is arguably the most widely used math education site on the Internet (search for “math” on Google.) It began in January of 1996 as a proof-of-concept grant from the NSF to extend the work of the Geometry Forum into other areas of mathematics and to investigate the viability of a virtual center for mathematics education on the Internet. The Math Forum has developed a vast Web site of over 925,035 learning resources and it receives over 650,000 visitors a month, with mentored user services such as Ask Dr. Math, for students of all ages, Problems of the Week services for grades 3-12, and Teacher2Teacher for discussions of pedagogy.

The Math Forum’s home page allows browsing and searching the Internet Mathematics Library of over 8600 annotated entries of hand-selected resources. The cataloguing features are based on American Mathematical Society categories, and are enhanced by recommendations of the American Mathematics Metadata Task Force.

The Math Forum provides many ways for people to interact with one another, with different points of access for people of varied strengths, needs, and interests. Community building is an important part of Forum activities and has formed the basis of much of the content development on the site.
Proposals for Research

The Math Forum represents a vision about the possibilities for an Internet community that extends the collegiality found in schools, classrooms, or the workplace. Evaluation of the Forum is used in program design, development, and facilitation, and provides an assessment of impact.

**JOMA Applet Project**

*DLI-2 Award Number 9980185*

Its goals were to 1) search the Web and other resources to locate and collect applets and similar programs developed by the mathematics research and teaching Communities, 2) review and test these systematically, and 3) to make them easily accessible to undergraduate faculty and students. *JOMA, the Journal of Online Mathematics and its Applications,* is published by the Mathematical Association of America. This project was the basis for MathDL an undergraduate-level digital library, NSDL Award Number 0085861, a joint project between the MAA and the Math Forum, which is developing the technical infrastructure.

These projects have given us considerable experience constructing libraries and supporting technologies, such as metadata for the NSF digital library initiative. In addition, numerous Forum staff members have contributed to NSDL activities, meetings and working groups. The Math Forum was a founding member of the SMETE Open Federation, the largest identifiable user base for the National STEM Education Digital Library.

**ESCOT (Educational Software Components of Tomorrow)**

*REC Award Number 9804930*

This was a testbed for the integration of innovative technology in middle school mathematics. The Math Forum, working with SRI and other partners, developed team-based approaches that produced math tools for integration into the Problems of the Week.

**The Math Forum Online Mentoring Project**

*DUE Award Number 0127516*

This is developing a guide to enable professors to integrate online mentoring experiences into their mathematics and mathematics education courses. Pre-service teachers in these courses mentor students submitting their solutions to the Math Forum's Problems of the Week. The results of this project will...
be used to train mentors for the Technology Problem of the Week (tPOWs), part of a new NSDL funded digital library of mathematics software.

Organizational Memory and Organizational Learning (CSS)

“Conceptual Frameworks and Computational Support for Organizational Memories and Organizational Learning (OMOL),” PIs: Gerhard Fischer, Gerry Stahl, Jonathan Ostwald, September 1997 – August 2000, $725,000, from NSF CSS Program #IRR-9711951.

This grant was instrumental in the PI’s turn from earlier work on organizational memory to support for collaborative learning. The OMOL project started from a model of computer support for organizations as Domain-Oriented Design Environments (DODEs) in which both domain knowledge and local knowledge are stored in the form of artifact designs and associated design rationale (Fischer, 1994). This CSCW model evolved into one of Collaborative Information Environments (CIEs), that emphasized the interactive, asynchronous, persistent discussion of concepts and issues within an organization (Stahl, 1998; 2000a). Gradually, interest in organizational learning aspects led to involvement in CSCL and the model of collaborative knowledge-building environments (Fischer et al., 1999). A number of software prototypes were developed to explore the use of the Web as a communication and collaboration medium. Of these, the most important for the proposed work was WebGuide a prototype threaded discussion system that provided multiple perspectives on the discussion, comparison of perspectives and control over rearrangement of notes (Stahl & dePaula, 1998; Stahl et al., 1998). Deployment of WebGuide in classrooms raised serious issues of adoption and concerns of socio-technical and social informatics (Kling, 1999) issues: motivation, media competition, critical mass, social practices, seeding, management, re-seeding, convergence of ideas, peer-to-peer collaboration, deployment strategies.

WebGuide and Environmental Perspectives (NOAA)


This grant funded the initial implementation of WebGuide as an integrated Java applet supporting personal and group perspectives. It was a joint effort between the PI, a middle school teacher, and a research group at the NOAA labs in Boulder. The teacher taught an environmental science class in which he wanted to spend the year having his students interview various adults and
construct a set of contrasting perspectives (conservationist, regulatory, business, community) on a particular local environmental issue that the students had previously been involved in. WebGuide was used by the students to collect notes on their interviews and to formulate personal and team perspectives on the issue. Results of this software trial were analyzed and presented at conferences (Stahl, 1999a; 1999b; 1999c; Stahl & Herrmann, 1999).

Innovative Technology for Collaborative Learning (European Commission)


This grant supported software design and development of BSCL by researchers in Finland, Germany and Spain. The software was implemented as extensions of BSCW, a mature CSCW product used by 200,000 unique users since 1996 (Appelt, 1999). The PI went to work with the BSCW team at Fraunhofer-FIT near Bonn, Germany, for the first year of the project. He prototyped the BSCL innovations and published descriptions of them (Stahl, 2002d; Stahl, 2003b). During its second year, the project is assessing the use of the new software in schools in Finland, Netherlands, Italy and Greece.

8. International Collaborations

The proposed NSF Project builds on the work of the European ITCOLE Project and its BSCL software. The PI was the primary designer and prototyper of the BSCL software when he worked at Fraunhofer-FIT in Germany. The Project with the Math Forum will involve close collaboration with the BSCW/BSCL team at FIT and has their full support. FIT will continue to support the BSCL code, making it available for free to educational institutions throughout the world. They will also provide training to Project staff who will be modifying the BSCL code. FIT has granted a five year developers license to the PI to work on extending BSCL as part of this Project. Both Wolfgang Appelt, the BSCW/BSCL team manager and Wolfgang Prinz, the director of the CSCW department at FIT personally support the proposed Project and its collaboration with FIT (see Supplementary Documentation).

The idea of automated support for group formation for workgroups in online learning is a research topic at the Distance University of Germany (Fern-
Uni, Hagen). Joerg Haake, who has begun research on this topic (Haake et al., 2003; Wessner et al., 2002) will be a close collaborator with this Project.

In general, a number of leading international HCI, CSCW and CSCL researchers have already agreed to collaborate on this Project, participating in the workgroups that will conduct much of the project planning, experimental designing, software design and assessment. Others will be added as needed. The names, affiliations and research interests of these international collaborators are listed in the attached Biographical Sketches section.

9. Project Assumptions, Hypotheses & Methods

The global design of the Project is diagrammed in Figure 1. The core question is how best to balance group formation and how best to support the collaborative learning of the formed groups. The global hypothesis is that carefully balanced groups and properly supported or coached groups will learn better and produce higher quality results. This hypothesis is tested by comparing groups in which differences among the members have been balanced in accordance with various algorithms with randomly assigned or self-selected groups. Then, an assortment of software tools and structures will be used to support the collaboration processes of the groups. The results of groups with these supports will be compared with the results of groups without these special supports. Thus, the Project will not only demonstrate that groups in the upper right quadrant learn better (assuming the global hypothesis is confirmed), but it will also be able to distinguish the effects of the structure of the catalyzed groups from the effects of their nurturing.

To test the global hypothesis, a number of assumptions and secondary hypotheses are made:

Assumption 1: Collaborative learning is good. While we will investigate the processes of collaborative learning and discover much about its power, its difficulties and its limitations, this Project will not focus on comparing it with individual learning.

Assumption 2: Small groups are useful for building communities. The Project looks at virtual communities like the Math Forum as being built out of the interactions of individual members as they participate in small groups; it will observe the forms of leadership, changing roles, group continuities and community structures that emerge and evolve as groups are catalyzed or as they spontaneously form.
Assumption 3: It is useful and sufficient to focus on the group as the unit of analysis. While it is necessary to analyze the contributions of individuals to a group’s learning, it is possible to interpret the meaning of group discourse without making assumptions about or investigating the psychological states or personal interpretations of the individual members. The Project will only use instruments like individual interviews and post-tests in focused and limited ways.

Assumption 4: The success of collaborative learning at the group level can be measured by quantitative evaluation of the group products (such as solutions to math problems and the rationale for the solution). In addition, the Project will take a qualitative look at group process, but this will not provide the primary judgment of a group’s success.

Assumption 5: The intertwining of differing perspectives is productive of group learning.

Hypothesis 1: Catalyzing carefully balanced heterogeneous groups will create groups that learn better than randomly assigned groups.

Assumption 6: Collaboration is a complex process and people must learn how to interact productively.

Hypothesis 2: Providing tools that help groups to structure their interactions, that support specific phases of their collaborations and that coach their group process will create groups that learn better than they would have without those tools.

Method 1: The Project will assess the effect of balanced formation and support for collaboration primarily by quantitative evaluation of the group’s produced solutions to problems (see Figure 2).
Method 2: Secondarily, the Project will analyze the group interaction to assess the quality of the collaborative process.

10. Plan of Work

The five year Project period is planned to be January 1, 2004 – December 31, 2008. Roughly, work during these years will be focused as follows: Year I (2004) pilot studies; Year II (2005) controlled experiments; Year III (2006) prototype evaluations; Year IV (2007) field studies; Year V (2008) assessment and dissemination.

This Project is based on a belief in the power of collaborative knowledge building. Therefore, the Project will be a highly collaborative effort – and the success of that effort will itself be an object of study in the Project. Accordingly, the work will be conducted through the organization of three sets of workgroups: student workgroups, curricular workgroups and infrastructure workgroups. Project staff will be responsible for organizing these workgroups and coordinating the activities and results of the Project.
Project Staff

Project staff is based at Drexel University and works out of the Math Forum offices. Their primary activities will be related to supporting the catalyzing and nurturing of the student, curricular and infrastructure workgroups. Staff will be responsible for ensuring that these groups get started and function effectively to carry out their roles in the Project. It is acknowledged that most of the workgroup members are volunteers with significant limits on their time and that staff will have to support them to allow them to focus their effort effectively on making their most important contributions. Staff will work closely with these groups (often as members of the curricular and infrastructure workgroups) and will carry out the daily activities needed to implement the plans of these groups, such as drafting detailed proposals for the groups to review, transcribing videotapes, prototyping / implementing / releasing newly designed software, conducting statistical analyses of data, collecting / documenting / preserving Project data and materials. They will make all Project data and materials available online to the infrastructure workgroups and will ensure that these are preserved for future purposes. They will ensure that all logs and transcripts are distributed and preserved in a manner that protects privacy of students and guarantees anonymity. Project staff will be responsible for the successful functioning of the Project and for the production of required reports and other documentation.

Student Workgroups

Student workgroups are primarily organized online, although some Year I activities will take place face-to-face in local school classrooms. Student workgroups will increasingly be supported by online tools created as part of the project. They will be one of the sets of workgroups studied in the Project.

Year I: Pilot studies will explore collaborative learning using Math Forum problems in local middle- or high-school contexts, in face-to-face situations or with commercial mediation systems like instant messaging. Groups of 3 to 6 students will be selected by hand, with teacher input. Principles of group selection and group process from the research literature and from project hypotheses will be tested in this context. F2F interactions will be videotaped, digitized and transcribed. Transcripts and computer logs will be logged and reviewed to identify key interactions, which will be subjected to detailed discourse analysis and complementary forms of micro-ethnographic analysis. This is a highly explorative, qualitative phase of the project to build a focused experience base of small student groups working on Problems of the Week.
Year II: Controlled experiments of groups working on Math Forum problems will compare groups with different selection criteria: self-selection, teacher selection, homogeneous matching, heterogeneous balancing, knowledge complementing, random. These experiments will generate quantitative data related to evaluation of the group’s problem solution. Experiments will also be conducted to explore different kinds of math problems: longer time periods, more open-ended, more discussion-oriented.

Year III: Prototype evaluations will start to introduce tools for group selection and group scaffolding designed as part of the Project and based on the results of the previous two years’ results. Experiments this year will focus on alternative forms of scaffolding, including automated group process guidance, human mentoring, tools to support specific group processes (brainstorming, discourse, knowledge management, math representations, negotiation, textual references).

Year IV: Field studies will be done by enhancing the Math Forum environment with tools based on previous results. This will permit students from the broader Math Forum virtual community to be invited into workgroups. The interaction of selection criteria and forms of scaffolding will be studied (e.g., whether certain tools are more helpful for certain types of groups).

Year V: Experiments in the final year are reserved for unanticipated follow-up studies whose need is indicated by preliminary assessment of the Project results. This is a time to scale-up use of the group formation and scaffolding tools to become a normal part of the Math Forum. Observations will be made of the impact of collaborative student workgroups on the growth and dynamics of the virtual community and on the community’s impact on public interest in mathematics.

Curricular Workgroups

Curricular workgroups are primarily organized online, although some activities will take place face-to-face at the Math Forum or at special workshops held annually for mentors, teachers and student teachers in the curricular workgroups. Curricular workgroups will increasingly be supported by online tools created as part of the project. They will be one of the sets of workgroups studied in the Project.

Year I: Pilot studies will explore collaborative design of new Math Forum problems, in face-to-face situations or with commercial mediation systems like instant messaging. Problems will be designed for use by collaborative
workgroups of students. Groups of 3 to 6 participants will be selected by hand, from the pool of teachers, student teachers and mentors who assist in defining new math problems. Principles of group selection and group process from the research literature and from project hypotheses will be tested in this context. F2F interactions will be videotaped, digitized and transcribed. Transcripts and computer logs will be logged and reviewed to identify key interactions, which will be subjected to detailed discourse analysis and complementary forms of micro-ethnographic analysis. This is a highly explorative, qualitative phase of the project to build a focused experience base of small adult groups devising Problems of the Week.

Year II: Controlled experiments of groups designing Math Forum problems will compare groups with different selection criteria: self-selection, staff selection, homogeneous matching, heterogeneous balancing, knowledge complementing, random. These experiments will generate quantitative data related to evaluation of the group’s problem generating ability. In addition to designing new problems suitable for collaborative student workgroups, the curricular workgroups will design ways of incorporating such collaborative problem-solving in school classrooms, possibly mixing workgroups across schools or even across countries (taking into account issues of language and time zones, for instance).

Year III: Prototype evaluations will start to introduce tools for group selection and group scaffolding designed as part of the Project and based on the results of the previous two years’ results. Experiments this year will focus on alternative forms of scaffolding, including automated group process guidance, human mentoring, tools to support specific group processes (brainstorming, discourse, knowledge management, math representations, negotiation, textual references). The tools for adults searching the problem database and designing new problems may be different from the tools for students solving problems.

Year IV: Field studies will be done by enhancing the Math Forum environment with tools based on previous results. This will permit teachers and mentors from the broader Math Forum virtual community to be invited into these workgroups as they demonstrate interest and background for this work. The interaction of selection criteria and forms of scaffolding will be studied (e.g., whether certain tools are more helpful for certain types of groups).

Year V: Experiments in the final year are reserved for unanticipated follow-up studies whose need is indicated by preliminary assessment of the Project results. This is a time to scale-up use of the group formation and scaffolding tools to become a normal part of the Math Forum. Observations will be
made of the impact of curricular mentor workgroups on the growth and dynamics of the virtual community, such as how it enables adults to move into more central forms of community participation.

**Infrastructure Workgroups**

Infrastructure workgroups consist of Project staff and collaborators from the international CSCL community. They are primarily organized online, although some activities will take place face-to-face at annual workshops associated with CSCL conferences. Infrastructure workgroups will increasingly be supported by online tools created as part of the project. They will be one of the sets of workgroups studied in the Project. The infrastructure workgroups will help to clarify applicable theory, point to existing relevant literature, define appropriate pedagogy, develop experimental methodology, specify technology requirements, evaluate software designs and provide on-going formative assessment of Project progress.

Year I: Four infrastructure workgroups will be formed, each including European researchers, American researchers and Project staff. Each group will have a multidisciplinary mix of skills: technical, psychological and pedagogical. They will design and monitor the pilot studies of the student and curricular workgroups and will participate in the analysis of the results.

Year II: The infrastructure workgroups will design and monitor the controlled experiments of the student and curricular workgroups and will participate in the analysis of the results.

Year III: The infrastructure workgroups will design and monitor the prototype evaluations of the student and curricular workgroups and will participate in the analysis of the results. They will also start to use some of the new tools for their own collaborations.

Year IV: The infrastructure workgroups will design and monitor the Field studies of the student and curricular workgroups and will participate in the analysis of the results, including studying the interaction of selection criteria and forms of scaffolding.

Year V: In Year V, the infrastructure workgroup members will disseminate the results of this Project at conferences, other projects and publications, both in the US and abroad. They will analyze the effect of active workgroups on the growth and dynamics of the larger virtual community, such as how it results in community knowledge and in changing patterns of participation.
11. Anticipated Results & Impact

Intellectual Merit: This Project explores a primary open challenge of the Internet – with detailed and rigorous methods, under controlled and real-world global conditions: how to foster effective collaborative online learning. It joins the multidisciplinary expertise of the international CSCL community with the practical success of the Math Forum to study how to mediate the growth of a large virtual learning community, and to design, develop and assess tools for the automated support of small workgroups acquiring, managing and negotiating knowledge.

Innovation in IT: An unfulfilled promise of the Internet is to bring together people who do not know each other or live close by, but who could benefit from interacting within knowledge-rich contexts. This Project addresses core issues of computer support for collaborative learning (CSCL): how best to form and structure intimate learning workgroups within global knowledge-building communities and how to effectively scaffold their interactions.

Integration of Research & Education: The Math Forum is a major practical success of prior NSF research, forming a virtual community of about a million students, teachers and mathematicians. This Project will greatly expand on-going efforts to form small collaborations within the large body of users who now interact with the site and learn mathematics primarily as individuals. This fundamental research into innovative support for small group collaborative online learning will take place within a vibrant and realistic large-scale context and will impact all levels: student motivation and learning, teacher development, and community evolution – generating a new model of global virtual learning communities, incorporating the power and motivation of small-group collaboration.

Broader Impacts: The Math Forum model, with automated formation of small groups and support for interactions developing deep understanding of mathematics will be suggestive for virtual learning communities in other domains, taking advantage of other digital libraries. This model provides opportunities for students and teachers excluded from vital collaborative learning due to geographic isolation, disadvantaged schools, physical disability, discrimination and other physical or social factors. The model stimulates both student motivation and teacher development. It fosters interest in mathematics by transforming it from a stigma into a bridge to global friendships.

Integrating Diversity: A central Project hypothesis is that groups integrating diversity of all kinds learn better.
International Collaboration: The Project builds on the PI’s prior work on an EU grant, including its pedagogy, software and assessment. Core aspects of the Project – including technology design, pedagogy and assessment – will be conducted by workgroups of American and European leaders of the CSCL community in collaboration with Project staff. Annual workshops at international conferences will bring these collaborators together with each other and with wider international audiences.

References


NSDL: Collaboration Services for the Math Forum Digital Library

Project Summary

Statement of need. NSDL is intended to serve learners in both collaborative and individual settings, as well as formal and informal modes. If one carefully studies learning in school, workplace and home, one finds that most learning is a subtle mix of collaborative and individual effort. Unfortunately, to date digital library services focus almost exclusively on the needs of individual users. Support for “collaboration” has been largely limited to mechanisms for anonymous, asynchronous collaboration within the whole user community, where results obtained by individuals may be fed back into metadata for future use by all. Little support has been developed for direct collaborative use of digital libraries by small groups of people working together.

Project approach. The adaptation of groupware components from current CSCW and CSCL systems makes it feasible to develop collaborative learning environments as digital library services, significantly increasing the potential impact, efficiency and value of digital libraries. This Project provides a model and test case of such an approach – within the successful Math Forum Digital Library (MFDL).

Target audience. The MFDL offers a variety of theoretical, practical, pedagogical, interactive and fun resources and services related to K-12 mathematics. It already supports a user community of close to a million distinct users. A popular service is the Problem of the Week (PoW), which is solved in and out of schools, by individuals and small groups. The MFDL now aims to extend the appeal and mathematical depth of these PoWs by bringing students together in small, online groups for asynchronous and synchronous collaborative learning at a distance.

The Project goals for advancing collaborative services in the NSDL are the following:
To better understand the computer support needs of small groups collaborating in a digital library.

- To design a collaborative learning environment within a digital library.
- To evaluate the use of a collaborative learning environment within a digital library.
- To incorporate a collaborative learning environment within a digital library as a sustainable service.

The *Project objectives* are to achieve these goals using the MFDL as a model and test case:

- To study the computer support needs of small groups of students (user teams) collaborating on PoWs in the MFDL.
- To develop special PoWs and associated curricular resources for collaborative usage, with the help of teachers and student teachers (creator teams). These teams will mine the MFDL and provide new resources to it as well as rate, annotate and organize existing resources.
- To design a Math Forum Collaborative Learning Environment (the MFCLE) within the MFDL, with the help of international CSCL (computer support for collaborative learning) researchers and developers (design teams).
- To prototype, evaluate and iterate the design of the MFCLE, in accordance with HCI best practices.
- To implement a stable version of the MFCLE, providing collaborative work areas and tools to communicate and collaborate with team members and other MFDL community members.
- To evaluate the use of the MFCLE by user teams, creator teams and design teams.
- To incorporate the MFCLE as a sustainable service of the MFDL.
- To disseminate the MFCLE as a reproducible model of a digital library service that promotes collaborative learning.

The *Project team* consists of four co-PIs, creator teams (student teachers, teachers and MFDL staff) and design teams (national and international CSCL researchers and MFDL staff). PI Stahl has developed numerous collaborative learning environments, has published on CSCW and CSCL theory, methodology and evaluation, and teaches HCI. Co-PI Weimar has been Director of the MFDL since its founding in 1994. Co-PI Bach is
Proposals for Research

professor of educational technology. Co-PI Shumar is an educational ethnographer and long-time evaluator of the MFDL.

**Intellectual merit.** This Project creatively combines leading-edge collaboration technologies with one of the most popular services of a successful digital library to provide a model of support services for collaborative digital library usage. The Project brings together four co-PIs with the required mix of expertise, along with teams of engaged educators and international researchers.

**Broader impact.** The Project develops collaboration services for digital libraries, providing a sustainable model. It promotes the involvement of geographically isolated, disadvantaged and disabled students, distributed teachers and international researchers by inviting them into collaborative learning teams hosted, supported and informed by a digital library. It pioneers a path for enhancing NSDL impact and building virtual learning communities.

**Project Description**

**1. Statement of Need**

1.1. **The Need for Collaboration Support**

NSDL is intended to serve learners in both collaborative and individual settings, as well as formal and informal modes. If one carefully studies learning in school, workplace and home, one finds that most learning is a subtle mix of collaborative and individual effort (Fischer & Granoo, 1995; Stahl, 2002d). Unfortunately, to date digital library services focus almost exclusively on the needs of individual users. Support for “collaboration” has been largely limited to mechanisms for anonymous, asynchronous collaboration within the whole user community, where results obtained by individuals may be fed back into metadata for future use by all. Little support has been developed explicitly for direct collaborative use of digital libraries by small groups of people working and learning together. This Project proposes to develop collaboration support technology for small groups pursuing a shared goal as a digital library service, and to demonstrate its effectiveness in the learning of school mathematics.

How will these small groups of people with a shared interest and/or complementary skills come together? The long-heralded promise of the
Internet was that people could not only access the whole world of information from any location, but also that they could meet people who shared their interests or could help them and that they could then explore ideas together (Bush, 1945; Engelbart, 1995; Hiltz & Turoff, 1978; Rheingold, 1993). With the exponential growth and the specialization of the world’s knowledge in mathematics and science, for instance, it is unlikely that someone with a particular momentary interest would happen to know many other people in their physical neighborhood with that interest at that time. Yet, there are currently few examples of technologies that fulfill the promise of the Internet and allow one to find and work with some of the many people who might share one’s interest across the nation or around the world.

Of course, we see with Internet newsgroups, for instance, that people already do struggle to engage in collaborative problem-solving even with the primitive nature of available tools and the haphazard character of opportunities for group formation. This Project hypothesizes that systematic support for the building of small groups will foster more effective collaborative learning, and that providing appropriate tools for their group interactions will further increase their effectiveness in taking advantage of digital libraries. It should also result in longer-term, more in-depth collaborative learning.

1.2. Target Audience: The Math Forum Digital Library Community

The Math Forum Digital Library (MFDL) – begun with NSF support – is an extensive mathematics website supported by a professionally staffed organization. It offers a variety of services, primarily to students and teachers interested in topics of mathematics commonly encountered from kindergarten through calculus courses. The digital library includes over a million web pages with FAQs, math challenges, discussions, interactive applets, articles and technical sources. It currently receives over 800,000 distinct online visitors a month. Most of these are people who are solving mathematics problems on an individual basis, whether from school, work or home.

A popular service of the MFDL is its “Problem of the Week” (PoW), which is solved in and out of schools, by individuals and small groups in classrooms. PoWs offer motivating opportunities for inquiry-driven learning that is active and engaging. The MFDL now aims to extend the appeal and mathematical depth of these PoWs by bringing students together in small, online groups (called “user teams” in this Project) for asynchronous and synchronous collaborative learning at a distance. MFDL is also interested in
supporting collaborative approaches for teachers developing new PoW curriculum (“creator teams”) and for people associated with MFDL who are developing new collaboration services (“design teams”). In general, MFDL would like to harness, extend and apply collaboration technologies to foster and support small groups of people within the MFDL virtual community.

Computer support for collaborative learning (CSCL) is an established research field, offering technology, theory and pedagogy that can be adapted, extended and applied in digital libraries (Stahl, 2002f). Although there are comprehensive CSCL systems to support classroom learning (e.g., WebCT, Blackboard, Lotus LearningSpace, Knowledge Forum / CSILE (Scardamalia & Bereiter, 1996), WISE / KEY (Slotta & Linn, 2000), Synergeia / BSCL (Stahl, 2002e)), these systems do not include support for the formation of small workgroups based on criteria of similarity of interest and complementarity of skills – even within classrooms, let alone in larger, more amorphous communities. The proposed Project will develop ways to support the formation and subsequent collaborative learning processes of a variety of specific types of small groups drawn from the MFDL community by taking advantage of what is known in CSCL.

This Project will approach group formation in a systematic way, developing theory, technology and pedagogy that are integrated together. The collaborative learning theory will describe the phases of group interaction, such as: group formation, task specification, brainstorming, proposals, negotiation and publication of results (Stahl, 2000c). The Project will build not only upon insights from the CSCL literature, but specifically on past studies of the MFDL virtual community (Renninger, Weimar, & Klotz, 1989; Renninger & Shumar, 2002a). The computer support technology will provide a comprehensive environment within which virtual groups can successfully pass though these phases (Stahl, 2002e). It will be developed using best practices of user-centered human-computer interaction (HCI) and iterations of user testing within MFDL (Preece, Rogers, & Sharp, 2002). The pedagogy will describe the nature of appropriate group membership criteria, problem characteristics and process facilitation. It will involve a reorganization of the usual MFDL process oriented to individual learning into one oriented to collaborative group learning, as described in this proposal.

The MFDL community can be described as a pyramid, with a broad base of individual first-time visitors, followed by successive layers of: loyal readers, contributors, facilitators and finally a small staff. While MFDL began in 1996 as a top-down, funded effort, its goal is to create a self-sustaining community where ideas, PoWs, and activities flow up from the bottom. The
formation of virtual groups will be a major means for achieving this, by encouraging and supporting people to move up the pyramid from occasional visitor to co-designer of MFDL services. In particular, groups will be formed at three levels in this Project:

1. User teams: visitors/readers who want to solve collaborative PoWs. They will work on problems that are particularly suited to collaborative inquiry, stressing richer, more open-ended topics designed to foster discussion of deep mathematical understanding.

2. Creator teams: contributors/facilitators who formulate the new collaborative PoWs. These groups of contributors will have access to the extensive digital library of past MFDL problems, math-related applets and studies of responses to previous PoWs. Collaborative approaches at this level of the pyramid are particularly appropriate for developing curriculum to be shared among teachers of a given grade, lifelong learners with a common special interest or mentors who want to answer questions collaboratively that are too time-consuming for individual mentors to answer.

3. Design teams: facilitators/staff who maintain and extend the MFDL technical infrastructure. Groups of researchers, programmers and MFDL staff will form to extend the MFDL through the production of new mathlets (programmable applets for computers and handhelds) and collaboration tools in response to the needs of the user and creator teams.

The Project is designed to be iterative, so that the different aspects of the Project can evolve in response to each other. The Project starts on the basis of considerable experience with PoWs solved by individuals – generally working within supportive communities of classrooms or MFDL mentoring, but without collaborative learning groups. The early Project phases focus on micro-analysis of this experience, incorporating previous studies of the MFDL virtual community, but clarifying the theoretical and pedagogical issues as well as the specific user requirements for technology. Then technological supports are gradually introduced to support group formation and collaborative knowledge building. An initial set of collaborative PoWs that have been meanwhile adapted from old PoWs is used. Later in the Project, groups at all three levels are active; well-defined user testing and formative evaluations drive revision of the technologies and practices. In the final phases, the Project will observe the stable functioning of self-forming groups at all three levels and will evaluate the success of people finding compatible group partners and learning mathematics collaboratively. Logging of group interactions as well as MFDL hits will allow for careful
quantitative and qualitative evaluation of the theory, technology and pedagogy.

1.3. Vision Scenario

User Team

It is January 2006 and the Project has just ended. Tanja is home schooled in up-state New York because of a physical disability; Sarah lives on a remote Navajo reservation; Damir attends school in Croatia. They each read the same collaborative PoW in the MFDL and became interested in it. The problem is to specify a general formula or algorithm for saying how many squares a straight line segment on graph paper will go through, given the end coordinates of the line. Based on previous visits of Tanja, Sarah and Damir individually, the MFDL website software determines that they have a mix of interests and skill levels that might work well in a small group solving this problem together. MFDL invites them and a couple more students to work together. Tanja, Sarah and Damir respond and find themselves together in MFCLE (the MFDL’s online collaborative learning environment). This shared virtual work environment helps them to coalesce into a group to work on this problem; to communicate both synchronously and asynchronously; to represent the problem and its features; to store, reflect upon and reorganize their collaborative ideas; to negotiate a group response to the problem; to document how they arrived at their solution; to submit their solution to MFDL; to receive feedback; and to decide if they want to continue collaborating.

Creator Team

Sandra, a teacher who has used MFDL with her math classes for years, reads the solution submitted by Tanja, Sarah and Damir and starts to think about a related problem, which she posts to a MFDL discussion. A number of other teachers respond with interest, and the MFCLE invites them to work together with a MFDL staff person to develop this idea into a publishable PoW. They also use MFCLE to collaborate, reviewing several of the responses to the previous PoW and eventually releasing a new problem to the MFDL that asks how many regions are formed by connecting N points on a circle to each other.

Design Team

MFDL staff notice that a number of recent PoWs involve drawing simple line figures and counting features such as vertices and regions. They request MFCLE to set up a work group with several of their technical collaborators around the world who might be appropriate and available. Together, the
people who respond form a group, specify user requirements, brainstorm designs, develop prototypes, conduct user testing and develop a new tool for the MFCLE. This tool allows student groups to collaboratively sketch representations that help them visualize and communicate about features of 2D drawings.

*Back to the present.*

This Project will explore the potential of digital libraries to bring together small groups of people like the three groups described above, who share interests and skills that might allow them to learn collaboratively and to build knowledge together. By the start of the grant period, MFDL is projected to be serving about a million students per month. This means that by the end of the grant period it will be common to have a thousand math-oriented people accessing MFDL at the same time during peak usage. This is a rich pool for forming matched small groups at various levels of mathematical interest for collaborative learning.

1.4. Impact

It is clear from current learning theory that collaborative learning is an effective way for many people to learn, particularly people who tend to be left behind in classroom situations (Johnson & Johnson, 1989). It is also clear from MFDL experience that an approach that emphasizes discourse and inter-personal interaction helps to build deep understanding of mathematical principles – rather than simply exercising rote memorization – in keeping with contemporary pedagogical priorities (NCTM, 2000; Renninger & Shumar, 2002b). What is not clear is how software should be designed to match people who are using digital libraries based on their known interests and skills, or whether these people can be formed into effective groups for collaborative learning. There has been almost no relevant research about online group formation (rare exceptions include: Haake et al., 2003; Wessner et al., 2002; Wessner & Pfister, 2001). Even in face-to-face settings, there are many open questions about how to form effective learning groups and then how to structure their collaborative tasks (Stahl, 2000b). Finally, although a variety of software is available to support collaborative knowledge building (Stahl, 2002d), it has not been incorporated into digital libraries. This Project will review the theoretical, technological and pedagogical issues and will systematically design, implement and assess an integrated approach to foster the collaborative building of mathematical knowledge in a digital library.
The availability of groupware components from current CSCW and CSCL systems that can be adapted to specific needs makes it feasible to develop collaborative learning environments as digital library services, significantly increasing the potential impact, efficiency and value of digital libraries. Collaborative small groups of different kinds can help to grow and structure a digital library, while building an active, engaged virtual community around the library. This Project provides a model and test case of such an approach – within the popular MFDL.

2. Project Goals

2.1. Goals

The Project goals for advancing collaborative services in the NSDL are the following:

1. To better understand the computer support needs of small groups collaborating in a digital library.
2. To design a collaborative learning environment within a digital library.
3. To evaluate the use of a collaborative learning environment within a digital library.
4. To incorporate a collaborative learning environment within a digital library as a sustainable service.

2.2. Objectives

The Project objectives are to achieve these goals using the MFDL as a model and test case. Specifically, the popular PoW service will be extended to collaborative PoWs for collaborative solution within a virtual learning environment, by achieving the following Project objectives:

1. To study the computer support needs of small groups of students (user teams) collaborating on PoWs in the MFDL.
2. To develop special PoWs and associated curricular resources for collaborative usage, with the help of teachers and student teachers (creator teams). These teams will mine the MFDL and provide new resources to it as well as rate, annotate and organize existing resources.
3. To design a Math Forum Collaborative Learning Environment (the MFCLE) within the MFDL, with the help of international CSCL
(computer support for collaborative learning) researchers and developers (design teams).

4. To prototype, evaluate and iterate the design of the MFCLE, in accordance with HCI best practices.

5. To implement a stable version of the MFCLE, providing collaborative work areas and tools to communicate and collaborate with team members and other MFDL community members.

6. To evaluate the use of the MFCLE by user teams, creator teams and design teams.

7. To incorporate the MFCLE as a sustainable service of the MFDL.

8. To disseminate the MFCLE as a reproducible model of a digital library service that promotes collaborative learning.

3. Project Design

3.1 Theoretical Framework

Collaborative learning, as understood in this proposal, involves a focus on the group level of analysis. Of course, construction of knowledge by a group can also be seen as co-construction by the individuals in the group, and conversely, the building of knowledge by the group has direct implications for learning by the participants. However, it is also true that the group can produce knowledge that none of its members would have produced by themselves (Fischer & Granoo, 1995; Hatano & Inagaki, 1991). For instance, the meaning of things said in group discourse is often defined by the group interaction itself rather than simply by ideas in the minds of individuals (Mead, 1934/1962; Stahl, 2003c; Wittgenstein, 1953). Although this Project will also be concerned with the learning by individual participants, its focus will be on supporting the building of knowledge by small groups. This emphasis is consonant with recent theories from cognitive science, communication theory, anthropology, education and CSCL, such as situated action (Suchman, 1987), activity theory (Engeström, 1999), situated learning (Lave & Wenger, 1991) and distributed cognition (Hutchins, 1996).

Traditional learning theory assumes that learning happens entirely in the mind of the individual, that it can be led or facilitated by a teacher, that the content is to some extent irrelevant to the process (in that one can learn anything this way), and that this is a primarily cognitive act. But advances in
learning theory have led to very different assumptions about the learning process. First is the realization that learning is a gradual and long-term process; people are always learning whether it is part of some specific curriculum or not. As Dewey pointed out, the purpose of teaching and education is not just to help students learn specific facts, but to create opportunities for experiences that lead to productive forms of learning through student exploration (Dewey, 1938/1991). That means connecting new experiences with a person’s prior experience base and creating opportunities for educative new experiences. In other words, creating a social context where new experiences can lead to the moral, emotional and intellectual development of the person. Starting from Dewey or Vygotsky (Vygotsky, 1930/1978), one strand of learning theory has focused on the individual in social context and how that context creates opportunities for learning. While there have been tremendous advancements here in our understanding of learning, learning is still often thought of by many followers of this strand as primarily something that happens to the individual (even if in a social context), and something that is primarily cognitive. Others have developed a more socially-oriented perspective.

Recent work on situated learning (Lave, 1991; Lave & Wenger, 1991; Lave, 1996) activity theory (Engeström, 1999) and cultural theories of learning (Cole, 1996; Holland, Hutchins, & Kirsh, 2000) have moved thinking about learning from the individual to the intersubjective experience, and from the cognitive to the whole person, including tacit and affective dimensions of learning. This work has led to several important implications for understanding learning: First, and perhaps most importantly, learning is a social process that involves individual interest, membership in a community where others are learning and engaged in productive social practice. Second, that knowledge is built intersubjectively and shared among members of the community. Finally the constraints on learning and knowledge production are constraints that exist within the social system, its form of organization and patterns of interaction, rather than predominantly within the individual. These important theoretical realizations about learning have tremendous consequences for collaborative learning and CSCL.

The emphasis on the group unit has methodological implications. The analysis of what takes place in Project investigations of MFDL and MFCLE usage will rely heavily upon interaction analysis (Duranti, 1998; Garfinkel, 1967; Heritage, 1984; Jordan & Henderson, 1995; Sacks, 1992; Stahl, 2002a) and community ethnography (Renninger & Shumar, 2002). These analyses will study in quantitative and qualitative terms how small groups function to mediate the establishment of a larger knowledge-building community, and how groups engage in sequences of different kinds of
interactions to build their knowledge. Tentative theories about how this knowledge-building process takes place will be subjected to empirical study, feeding back into revised formulations of a theory of collaborative knowledge building (Stahl, 2000c, 2002b, 2003b).

While we know a great deal about the social nature of the learning process, many of our educational and digital library institutions continue to be structured in ways that implicitly assume learning is individual and a matter of transferring information from teacher to student. That sad fact means that many of the things students actually learn most effectively in school are patterns of resistance from peers and commercial lessons from the marketplace that appeal more to their social and collaborative form of learning (Shumar, 1997). While some schools have begun to implement collaborative forms of learning, there are real limitations due to the political context of local schools and the difficulty of organizing resources and groups that cluster together concentrations of expertise and appeal to individual interests. Digital libraries such as the MFDL create an important strategic opportunity to bring a more collaborative, learning-oriented community of practice to individuals who may be distributed geographically in different school and institutional sites. Therefore, a critical next step is to figure out how to make online groups self-forming, successful and self-replicating.

3.2. Pedagogical Framework

Mathematics is often thought of as the discipline of “the right answer.” In 1998-2001, a small group of teachers and MFDL staff became uncomfortable with this designation because it can interfere with efforts to help students express their mathematical thinking, learn from mistakes, experiment effectively, and pursue their mathematical interests. They asked, how can we transform the student's question “Am I right?” into “How can I develop confidence and judgment that I am on the right track when working on a problem?” and “How can I know that I am improving my mathematical problem-solving and communication skills?” They decided that engaging students in discourse about mathematics was the way to go.

Discourse can make thinking public and create an opportunity for the negotiation of meaning and agreement (Bauersfeld, 1995). At the same time, discourse within a supportive and trusted small group provides collective support for developing one's thinking, drawing it out through the interest, questions, probing, and ideas of others (Cobb, 1995; Krummheuer, 1995; Wood, 1995; Yackel, 1995), and discourse enables students to connect their own everyday language with the specialized language of mathematics.
Articulating what they know allows students to clarify their own understandings. Through discourse, a teacher can better grasp the mathematical needs of the class: what the students know, misconceptions they may have, and how these might have developed (Resnick, 1988). Teachers and students gain perspective on their own thoughts through the attempt to understand the thinking of others, in the process laying the foundation for a supportive learning community (Brown & Campione, 1994).

Within the mathematics education community there is strong interest in the use of discourse for teaching and learning mathematics (Atkins, 1999; NCTM, 2000; Schifter, 1996). The teacher's role is described in broad terms as facilitative, to include listening carefully to students, framing appropriate questions, and mediating competing perspectives. Students are expected to develop problem-solving skills: defining problems, formulating conjectures, and discussing the validity of solutions. Stigler and Hiebert report similar roles for teachers and students in mathematics classrooms in Japan, where mathematical discourse is an integral part of instruction (Stigler & Hiebert, 1998).

An effective way to foster domain-oriented discourse is to catalyze active small workgroups. In heterogeneous small groups, students are challenged to stretch and learn within their “zone of proximal development” “in collaboration with more capable peers” (Vygotsky, 1930/1978). At the same time, the mentoring experience is productive for the “more capable” peer’s learning by teaching – and these roles are likely to reverse in other situations when the group members have complementary strengths.

The building of deepened understanding and increased knowledge of mathematics takes place in motivational community contexts, such as classrooms and research fields (Lave, 1991; Lave, 1996). Interactions within small groups can mediate effectively between individuals and these larger communities, providing supportive settings and engaging activities (Wenger, 1998). Small groups can build knowledge (collaborative learning) that draws upon and may extend the community knowledge while making it available to the individual participants who contribute to the group knowledge. According to theories of situated learning (Lave & Wenger, 1991), changing patterns of participation in which individuals become progressively more involved are important features of community learning. We have already seen signs of this taking place in the MFDL virtual community in the documented example of Sonia and her son (Renninger & Shumar, 2002, p. 66 ff). The MFDL already exploits and supports collaborative mechanisms in the community, for instance by archiving PoW
user solutions in a structured and indexed format designed to optimize accessibility and pedagogic impact. This Project will investigate the effects of online collaborative math learning by extending the services of the MFDL and its growing community. It will explore the effect this has in drawing average or poorly motivated students into intellectual engagement, as well as involving students and teachers already excited about math in a larger community.

3.3. Project Team

The Project team consists of four co-PIs (in various schools of Drexel University), creator teams (student teachers, teachers and MFDL staff) and design teams (national and international researchers and MFDL staff).

College of Information Science & Technology

Drexel University has a long history of technology leadership as a former Institute of Technology, including being the first university to require entering undergraduates to have a PC and more recently being judged the “most wired” university according to Yahoo.

Drexel University’s College of Information Science and Technology is rated the #1 graduate school of library science information systems by US News and World Report (http://www.usnews.com/usnews/edu/grad/rankings/lib/brief/infsp3_brief.php). This interdisciplinary college offers online and campus-based undergraduate and graduate programs in computer science (e.g., HCI, databases, software engineering) and library science (including digital libraries).

PI Gerry Stahl is an Associate Professor in Drexel University’s College of Information Science and Technology. He brings a multidisciplinary background to the Project, with PhD dissertations in philosophy/social theory and computer science/AI (Stahl, 1975, 1993b). He has developed a series of collaboration support systems: Hermes (Stahl, 1993a), WebNet (Stahl, 2000a), WebGuide (Stahl, 1999a, 1999b; Stahl & Herrmann, 1999; Stahl, 2001), BSCL (Stahl, 2002e, 2003a), and other educational software: Teachers Curriculum Assistant (Stahl, Sumner, & Owen, 1995; Stahl, Sumner, & Repenning, 1995) and State-the-Essence (Kintsch et al., 2000; Stahl & dePaula, 2001).

Stahl specializes in CSCL research, having published on CSCL theory (Stahl, 1993a, 1998, 2000b, 2002b, 2003b, 2003c) and the use of discourse analysis as an assessment methodology (Stahl & Sanusi, 2001; Stahl, 2002a, 2002c, 2002d). He was Program Chair of CSCL 2002 and Editor of the CSCL 2002 Proceedings (Stahl, 2002f). He is Workshop Chair of CSCL
2003 and Communications Chair and founding Board member of the International Society for the Learning Sciences (ISLS) (http://isls.org). He teaches online and in-class courses on HCI, CSCL and CSCW at Drexel, using small group collaborative learning methods.

The Math Forum Digital Library

Co-PI Steven Weimar has directed the MFDL since 1994. The MFDL is hosted at Drexel University. The MFDL began in 1992 as the Geometry Forum at Swarthmore College, expanded to the MFDL in 1996. It was funded in its development by the National Science Foundation, but has become largely self-sustaining in its stable services. It has become one of the most successful applications of the Internet to education through the development of interactive services that bridge the higher education, K-12, and industry communities. These services form the basis for a digital library that generates high quality mathematical content, supports student learning, integrates the benefits of technology with education, and is used for teacher professional development and pre-service teacher education. The MFDL now comprises over 1.2 million pages of content, has over 2 million visits a month, receives up to 9,000 queries a month at its “Ask Dr. Math” expert service, and mentored over 27,000 students during the 2000-2001 school year through its “Problem of the Week” services. Among its current projects are two NSF grants, one focused on the use of online student mentoring programs in pre-service teacher education courses, and the other on the development of MathTools, a digital library for software in mathematics education from arithmetic to calculus.

Education & Ethnography

Drexel University has a School of Education and a Department of Culture & Communication, both of which are represented in this Project. Co-PI Craig Bach is a professor in the School of Education, where he explores the use of technology in education, having developed several hypermedia presentations of topics in mathematics and philosophy. Co-PI Wesley Shumar is a cultural anthropology professor in the Department of Culture & Communication, who specializes in educational anthropology and has conducted ethnographic studies of the MFDL for many years.

3.4. Prior Work

The Math Forum

REC-9618223, $971,300, March 1999 to February 29, 2000
The MFDL is arguably the most widely used math education site on the Internet (search for “math” on Google). It began in January of 1996 as a proof-of-concept grant from the NSF to extend the work of the Geometry Forum into other areas of mathematics and to investigate the viability of a virtual center for mathematics education on the Internet. The MFDL has developed a vast Web site (http://mathforum.org) of over a million learning resources and it received more than 650,000 distinct visitors a month (making 2 million visits) in 2001, with mentored user services such as Ask Dr. Math, for students of all ages, PoW services for grades 3-12, and Teacher2Teacher for discussions of pedagogy.

The MFDL home page allows browsing and searching the Internet Mathematics Library of over 8,600 annotated entries of hand-selected resources. The cataloguing features are based on American Mathematical Society categories, and are enhanced by recommendations of the American Mathematics Metadata Task Force (http://mathmetadata.org/).

The MFDL provides many ways for people to interact with one another, with different points of access for people of varied strengths, needs, and interests. Community building is an important part of MFDL activities and has formed the basis of much of the content development on the site. The MFDL represents a vision about the possibilities for an Internet community that extends the collegiality found in schools, classrooms, or the workplace. Evaluation of the MFDL is used in program design, development, and facilitation, and provides an assessment of impact.

*Publications:* Virtual communities (Renninger & Shumar, 2002a, 2002b; Shumar & Renninger, 2002); Problems of the Week (Renninger & Shumar, 1998; Renninger, Farra, & Feldman-Riordan, 2000); geometry interactions (Renninger et al., 1989).

**JOMA Applet Project**

*DLI-2 Award Number 9980185*

The goals of this Project were to (1) search the Web and other resources to locate and collect applets and similar programs developed by the mathematics research and teaching communities, (2) review and test these systematically, and (3) make them easily accessible to undergraduate faculty and students. JOMA, *the Journal of Online Mathematics and its Applications*, is published by the Mathematical Association of America. This Project was the basis for MathDL.

**Bridging Research and Practice**

*REC Award Number 9805289*
BRAP was a joint program with TERC and Michigan State University investigating the possibilities for multimedia articles to open more effective communication between researchers and teachers. The MFDL developed a collaborative process through which teachers designed and conducted research into the use of discourse in the math classroom. A video-paper was produced jointly with researchers that served as the focal point for an online conversation with the mathematics education community at large. See http://mathforum.org/brap/wrap.

MathDL

NSDL Award Number 0085861

MathDL is an undergraduate-level digital library, a joint Project between the MAA and the Math Forum, which is developing the technical infrastructure. The MathDL and previous projects have given the Math Forum considerable experience constructing libraries and supporting technologies, such as metadata for the NSF digital library initiative. In addition, numerous Forum staff members have contributed to NSDL activities, meetings and working groups. The Math Forum was a founding member of the SMETE Open Federation, the largest identifiable user base for the National STEM Education Digital Library.

ESCOT (Educational Software Components of Tomorrow)

REC Award Number 9804930

The ESCOT Project was a testbed for the integration of innovative technology in middle school mathematics. The Math Forum, working with SRI and other partners, developed team-based approaches that produced math tools for integration into the Problems of the Week.

The Math Forum Digital Library Online Mentoring Project

DUE Award Number 0127516

The Online Mentoring Project is developing a guide to enable professors to integrate online mentoring experiences into their mathematics and mathematics education courses. Pre-service teachers in these courses mentor students submitting their solutions to the MFDL's Problems of the Week. The results of this Project will be used to train mentors for “Technology PoWs,” part of a new NSDL funded digital library of mathematics software.

Organizational Memory and Organizational Learning (CSS)

“Conceptual Frameworks and Computational Support for Organizational Memories and Organizational Learning (OMOL),” PIs: Gerhard Fischer,
Gerry Stahl, Jonathan Ostwald, September 1997 – August 2000, $725,000, from NSF CSS Program #IRR-9711951

This grant was instrumental in the PI’s turn from earlier work on organizational memory to support for collaborative learning. The project started from a model of computer support for organizations as domain-oriented design environments in which both domain knowledge and local knowledge are stored in the form of artifact designs and associated design rationale (Fischer et al., 1993). This CSCW model evolved into one of Collaborative Information Environments, that emphasized the interactive, asynchronous, persistent discussion of concepts and issues within an organization (Stahl, 2000a). Gradually, interest in organizational learning aspects led to involvement in CSCL and a model of collaborative knowledge-building environments (Stahl, 2001). A number of software prototypes were developed to explore the use of the Web as a communication and collaboration medium. Of these, the most important for the proposed work was WebGuide, a prototype threaded discussion system that provided multiple perspectives on the discussion, comparison of perspectives and control over rearrangement of notes. Deployment of WebGuide in classrooms raised serious issues of adoption and concerns of socio-technical and social informatics (Kling, 1999) issues: motivation, media competition, critical mass, social practices, seeding, management, re-seeding, convergence of ideas, peer-to-peer collaboration, deployment strategies.

WebGuide and Environmental Perspectives (NOAA)


This grant funded the initial implementation of WebGuide as an integrated Java applet supporting personal and group perspectives. It was a joint effort between the PI, a middle school teacher, and a research group at the NOAA labs in Boulder. The teacher taught an environmental science class in which he wanted to spend the year having his students interview various adults and construct a set of contrasting perspectives (conservationist, regulatory, business, community) on a particular local environmental issue that the students had previously been involved in. WebGuide was used by the students to collect notes on their interviews and to formulate personal and team perspectives on the issue. Results of this software trial were analyzed and presented at conferences (Stahl, 1999a, 1999b, 1999c; Stahl & Herrmann, 1999).
Innovative Technology for Collaborative Learning (European Commission)


This grant supported software design and development of BSCL by researchers in Finland, Germany and Spain. The software was implemented as extensions of BSCW, a mature CSCW product used by 200,000 unique users since 1996 (Appelt, 1999). The PI went to work with the BSCW team at Fraunhofer-FIT near Bonn, Germany, for the first year of the Project. He prototyped the BSCL innovations and published descriptions of them (Stahl, 2002e, 2003a). During its second year, the Project is assessing the use of the new software in schools in Finland, Netherlands, Italy and Greece.

Our current related work and related proposals

MFDL staff periodically try out mechanisms to support small group collaboration on a small scale. They have provided chat services or encouraged face-to-face groups in classrooms to submit team responses to PoWs. These trials generally produce immediate interest from the community, indicating that systematic support for small groups could have dramatic results in stimulating participation in the MFDL and the associated community.

The PI is exploring small group formation approaches and innovative software functionality to support small group collaboration in online courses using digital libraries. Each of his HCI courses engages in user studies, software design and user testing of specific applications in this area.

The co-PIs of this proposal recognize that many research and technical issues related to this Project require careful research and technology innovation that go well beyond the scope of this Project. They have therefore submitted an NSF ITR proposal for innovative technology to form and support small groups and will submit a ROLE proposal for related research on collaborative learning by small groups. Particular co-PIs are also involved in other projects and proposals, including the PI’s participation in an NSDL proposal for small group knowledge construction in college classrooms and co-PI Shumar’s participation in another NSDL targeted-research track proposal. These related projects – if funded – would be complementary to the Project proposed here, but mutually independent. Although co-PI hours might have to be adjusted, there would be different Research Assistants and different goals, objectives and timetables. The present proposal aims to quickly establish a model of collaboration services
Proposals for Research

in a digital library, based on research and technology that is almost at hand. Parallel research and innovation efforts would allow that model to be refined and extended in the future.

3.5. Infrastructure Technology

This Project aims to adapt existing technologies as much as possible and to combine compatible software components into an integrated environment to support collaborative use of a digital library by small groups working together on the Internet, specifically to support the solving of collaborative PoWs using the MFDL. Useful components for supporting collaborative communication are available in various configurations and on different programming platforms. There are, for instance, search, document exchange, email, chat, threaded discussion and whiteboard components in CSCL systems and in Open Source libraries. While it may not be feasible to develop specialized intelligent interfaces like Ariadne (Twidale & Nichols, 1998b) within the scope of this Project, the primary advantage of recording and displaying processes like goal definition, problem reframing, query refinement and result processing are obtained in a general way with persistent chat and threaded discussion tools. The only major component that has to be designed from scratch for the Project is a group formation component.

There do not seem to be any group formation components currently available, although the idea is not unprecedented (Swanson, 1964; Twidale & Nichols, 1998a). Some organizations have explored systems for locating expertise within their staffs (Ackerman & McDonald, 1996); but the techniques for that do not transfer to the problem of finding people with matching interests using a digital library. There have been some experiments with social awareness, to display other people who are viewing the same web page at the same time (Graether & Prinz, 2001), but this hint is not enough to support group formation. A “group formation” project in Japan matched learning theories (Inaba et al., 2000; Supnithi et al., 1999), but not people. A prototype for group formation in Germany allowed students who knew each other to self-select groups (Haake, Schuemmer, & Haake, 2003; Wessner & Pfister, 2001; Wessner, Dawabi, & Haake, 2002), but this approach does not scale to large groups who do not know each other personally. A spin-off of this German research is being expanded and developed for distance education; the Project will collaborate with Jörg Haake and associates through the design teams (see section on International Collaboration). It will also collaborate with H. Ulrich Hoppe and Bonnie Nardi, who have both prominently argued for supporting small group

The PI began exploring support for group formation while teaching an online HCI (Human-Computer Interaction) course for graduate students at Drexel. His students studied the issue and came up with several low-fidelity prototypes that they subjected to user testing. The PI developed an automated grouping agent, which he uses to form work groups in subsequent courses. In both the student prototypes and the grouping agent, groups were formed based on specific criteria about the participants: their schedules, their interests and their skill levels. These pilot studies for the proposed Project suggest the kinds of balance that should be sought in forming distributed groups. For instance, if synchronous communication is to be possible within the group – especially given different global time zones – members must have similar schedules. On the other hand, collaborative teams often work best when there is a diversity of perspectives and skills, along with a commonality of interests. Thus, a matching algorithm must optimize certain similarities and other differences. Various theories of collaboration stress the power of heterogeneity, of the utility of seeing things differently: cognitive dissonance (Festinger, 1957), perspectives (Boland & Tenkasi, 1995; Goldnam-Segall, 1998; Stahl & Herrmann, 1999), interdependence (Johnson & Johnson, 1989), zone of proximal development (Vygotsky, 1930/1978), cognitive flexibility (Feltovich et al., 1996).

A pilot study of group formation was conducted by the PI with classes using two different online collaboration environments: Blackboard and BSCL. Blackboard is a commercial system to support collaboration. It is used widely in university courses, particularly in the US. Blackboard can be extended (in Java) by third party developers using the Blackboard Building Blocks SDK (see http://buildingblocks.blackboard.com/bin/bbdn_info.pl).

BSCL (Basic System for Collaborative Learning) is a system with collaboration support for classrooms that is similar to Blackboard (Stahl, 2002d). It was designed and developed by the PI and others in 2001/2002 as part of a European Union research Project. BSCL is an extension (developed in Python) to BSCW (Appelt & Klöckner, 1999), a shared repository CSCW system widely used in European research and learning organizations. It is available for free to academic organizations. The PI has a license to develop it during the period of this Project (see Letter of Support in Supplementary Documentation).

The MFDL already has an infrastructure of custom software (developed in an object-oriented Perl-based environment) to support the virtual
community and digital library of math resources and activities. It is possible to extend this system in various directions, such as using ZOPE or other Open Source components, extending Blackboard or adapting features of BSCL. Java applets can also be developed, adapting from the PI’s WebGuide system. The Project will select one of these approaches during its early phases.

4. Plan of Work

4.1. Timeline

The two year Project period is planned to be January 1, 2004 – December 31, 2005. Roughly, work during these years will be focused as follows, based on Drexel University’s quarter calendar. Here are the major software system development efforts for the Math Forum Collaborative Learning Environment (MFCLE) by quarter:

• Winter 2004 – Project start-up
• Spring 2004 – User studies of groups working on PoWs
• Summer 2004 – Explore multiple designs for the MFCLE
• Fall 2004 – Prototype an initial version of the MFCLE

• Winter 2005 – Test the prototype with user teams
• Spring 2005 – Develop a robust version of the MFCLE
• Summer 2005 – Debug & refine the MFCLE; integrate it into the MFDL
• Fall 2005 – Project wrap-up and dissemination

User teams will be formed throughout the Project to work on collaborative PoWs. They will use online collaboration technologies from early on, gradually adopting the MFCLE as it becomes available. Their work with these technologies will be studied to determine user requirements of the software in the first quarters and to evaluate the various versions of the software later.

Creator teams will develop collaborative PoWs throughout the Project for use by user teams and for adoption in the MFDL. Creator teams will also use online collaboration technologies throughout, gradually adopting
Proposals for Research

versions of the MFCLE as they becomes available in order to experience first hand the affordances of these environments.

Design teams will focus on design of the MFCLE technology, initially reviewing available components, then designing an integrated environment, and later evaluating it in user tests. The design teams will also use online collaboration technologies throughout, gradually adopting the MFCLE as it becomes available in order to experience first hand its affordances.

Project objectives will be achieved by meeting the following milestones:

1. June 2004  – Produce a user requirements document specifying the major components and functionality for MFCLE.
2. August 2004  – Produce at least 5 PoWs specifically designed for use by collaborative teams.
3. October 2004  – Produce at least 3 alternative designs for an initial version of the MFCLE.
4. February 2005  – Produce a working prototype of an initial version of the MFCLE capable of being tested by user teams.
5. May 2005  – Produce a formal evaluation of the prototype with user teams.
7. October 2005  – Incorporate the MFCLE into the MFDL.
8. December 2005  – Disseminate the MFCLE model by releasing MFCLE to the MFDL community, by submitting at least 3 papers in international conferences, and by sharing Project results with the NSDL community and with the international researchers involved in the Project.

4.2. Management Plan

The PI, Stahl, will have primary responsibility for all aspects of the Project. Weimar and Stahl will share Project fiscal management, with accounting maintained by the Math Forum and Drexel University. Weimar and Stahl will share Project staff management, recognizing that many staff are long-time employees of the Math Forum, contributing part-time.

The Project Management Team consists of the four co-PIs and will meet twice a month.
The Project Staff consists of the four PIs, four Math Forum curricular staff, three Math Forum technical staff and a Project graduate research assistant:

- G. Stahl, Information Science – Design Teams Coordinator
- S. Weimar, Math Forum – User Teams Coordinator
- C. Bach, Education – Creator Teams Coordinator
- W. Shumar, Anthropology – Evaluation Coordinator
- I. Underwood, Math Forum – MFDL Ask Dr. Math
- A. Fetter, Math Forum – MFDL Problem of the Week
- K. Lasher, Math Forum – MFDL Problem of the Week
- S. Alejandre, Math Forum – MFDL Problem of the Week
- L. Smith, Math Forum – MFDL IT director
- D. Tristano, Math Forum – MFDL software developer
- J. Zhu, Math Forum – MFDL system administrator
- GRA, Information Science – software developer

The Project Staff will hold monthly meetings at the Math Forum offices. These meetings will plan detailed Project milestones and activities; review progress made according to the milestones; prepare for up-coming activities; review and revise the Project plan; and make other decisions about the Project as needed. Minutes of these meetings will be posted on the Project website with other Project resources for review by the design teams, acting as Project advisors.

Project management will be conducted following a collaborative model, in keeping with the philosophy of the Project. Project activities will involve the collaborative teams, with Project staff providing staff support and taking responsibility to ensure tasks are accomplished. Each set of teams will be coordinated by a co-PI: Weimar (user teams), Bach (creator teams), Stahl (design teams). The Project takes an assessment-heavy approach to investigating the requirements for and effectiveness of technology; Shumar will coordinate the experimental design, ethnographic investigation, and formative and summative assessment of the Project. Stahl is responsible for software design and development; Smith for integration of software into the MFDL site; Weimar for involvement of students and teachers, as well as integration of Project activities with other MFDL activities; and Bach for pedagogical aspects of the Project.
Development of collaborative PoWs and other curricular materials will be done through the creator teams, consisting primarily of teachers and student teachers. The design teams – including national and international researchers as well as Project staff and interested members of the creator teams – will assist in the design and evaluation of Project experiments and of software for use in the experiments; they will monitor and guide the progress of the Project. All teams will be encouraged to be self-reflective and to become increasingly involved in the Project.

MFDL PoW staff will participate in planning, design and facilitation of the user and creator teams. MFDL staff will help with logistics, using their existing systems and networks of contacts. They will also help with hosting workshops for the teams as needed.

Shumar will coordinate all data collection, and will focus the teams as needed for formative evaluation tasks. Stahl is responsible for Project reports, including annual reports to NSF, culling from team summaries. Stahl, Bach and Shumar will prepare papers for conferences. Stahl and Weimar will be responsible for dissemination within the NSDL community.

4.3. National and International Collaborators

An important feature of this Project is the involvement of leading national and international researchers in the design teams. They bring expertise from a variety of relevant specialties and perspectives. Their participation will provide a natural means for sharing practical knowledge from Europe and the US as well as for disseminating the results of this Project across the nation and globe. To ensure a strong cadre of collaborators, the following researchers have already expressed strong interest in participating in the Project; others can join in the future:

National

Geri Gay (Cornell), Ricki Goldman-Segall (NJIT), Cindy Hmelo-Silver (Rutgers), Christopher Hoadley (Penn State), Timothy Koschmann (Southern Illinois U), Bonnie Nardi (Agilent), Leysia Palen (Colorado), Linda Puliam (California State U.), Mark Schlager (SRI), Dan Suthers (Hawaii).

International

Wolfgang Appelt (Fraunhofer-FIT, Germany), Thanasis Daradoumis (Barcelona, Spain), Hugo Fuks (Rio, Brazil), Jörg Haake (Distance U, Germany), Kai Hakkarainen (Helsinki, Finland), Thomas Herrmann (Dortmund, Germany), Ulrich Hoppe (Duisburg, Germany), Jim Hewitt
Proposals for Research

(Toronto, Canada), Victor Kaptelinin (Umea, Sweden), Anders Morch (Oslo, Norway), Wolfgang Prinz (Aachen, Germany), Barbara Wasson (Bergen, Norway), Volker Wulf (Siegen, Germany).

These individuals are established leaders in the HCI, CSCW and CSCL research communities, having made important contributions in theory, system design and assessment methodology. They all recognize the importance of collaboration, both in theory and in practice. See the Biographical Sketches section for more information.

The proposed NSF Project builds on the work of the European ITCOLE Project and its BSCL software. The PI was the primary designer and prototyper of the BSCL software when he worked at Fraunhofer-FIT in Germany. The Project with the MFDL will involve close collaboration with the BSCW/BSCL team at FIT and has their full support. FIT will continue to support the BSCL code, making it available for free to educational institutions throughout the world. They will also provide training to Project staff who will be modifying the BSCL code. FIT has granted a five year developers license to the PI to work on extending BSCL as part of this Project. Both Wolfgang Appelt, the BSCW/BSCL team manager, and Wolfgang Prinz, the director of the CSCW department at FIT, personally support the proposed Project and its collaboration with FIT (see Supplementary Documentation).

The idea of automated support for group formation for workgroups in online learning is a research topic at the Distance University of Germany (Fern-Uni, Hagen). Jörg Haake, who has begun research on this topic (Haake et al., 2003; Wessner et al., 2002) will be a close collaborator with this Project.

4.4. Project Evaluation

The Project will be considered successful if it achieves the objectives stated in Section 2.2 and meets the associated milestones stated in Section 4.1. But evaluation also plays two non-trivial roles in the work of this Project: (1) on-going testing of the software as an integral part of the user-centered design of the new technology, and (2) study of collaborative learning in a digital library as promised in Goals 1 and 3 as stated in Section 2.1. These two roles can be fulfilled by an ethnographic approach.

Evaluation for the Project is designed to provide specific data about the quality of interactions in the different kinds of teams using MFCLE. Data collected will largely be descriptive ethnographic data, which is appropriate to the needs of the Project. The goal will be to provide a detailed description of the interactions within each of the kinds of teams and to interview team
participants to capture their feelings about how well their groups worked. These descriptions will allow Project staff to assess which teams are doing well and which ones are less successful. Drawing on prior MFDL work with the ESCOT Project, teams will be evaluated in terms of their ability to communicate, develop a sense of shared worldview and create a feeling of group belonging – all of which contribute to successful work practices (Shumar, 2002; Lave & Wenger, 1991; Wenger, 1998). Analysis of studies of the teams will also contribute to the overall evaluation of the Project and the success of its implementation.

In year I the analysis of user teams will consist of two categories: face-to-face groups and virtual groups. We will observe two sites in schools where collaborative group work is ongoing. These may be groups using PoWs and they may be doing other projects. The face-to-face sites will involve extensive observation over the period of the collaborative problem solving. This may involve regular classroom participation for a week or two. Interactions will be videotaped and participant observation data will be collected. In addition to the two face-to-face sites, four virtual workgroup sites will be established. These will be virtual groups of students who have volunteered to work collaboratively on the math problems. Data from these groups will be collected on synchronous and asynchronous forms of interaction (chat transcripts, discussion lists, emails, and interviews with participants). Preliminary analysis of this student data will assess the patterns of interaction and begin to create a typology of successful group dynamics, as well as get participants’ sense of the quality of the group interaction. Drawing on earlier work on mathematical thinking at the MFDL, interactions will also be assessed for the quality of the work that went into the problem-solving in the group (Renninger, et al., 2000).

Face-to-face work in teams will be videotaped. The videos will be time-stamped and logged. Interesting episodes will be carefully transcribed. The MFCVE software will be instrumented to log usage data, including digital library queries submitted. Interactions captured will be coded at the utterance level, using grounded theory techniques to develop an appropriate coding scheme (Strauss & Corbin, 1998). Particularly rich interactions will be subjected to discourse analysis (Duranti, 1998; Jordan & Henderson, 1995; Sacks, 1992).

The year I creator team evaluation will focus on the analysis of two teams over the course of the year. Interactions in these teams will be tracked on synchronous and asynchronous forms of interactions (chat transcripts, discussion lists, email, and interviews). Face-to-face interactions of the teams will be videotaped and observed directly. Analysis of group
Proposals for Research

interaction and discourse will center around the emerging patterns of leadership, creation of a sense of group belonging, and the ability to communicate across differences of group, culture, need, etc. Finally, two design teams will be evaluated by collecting virtual chat interactions, discussion lists and interviews with members of the group. Analysis of group interaction will follow the same pattern as the curricular workgroups.

In year II, evaluation of the user teams will follow a similar format to the evaluation of the virtual groups in year I. Five groups will be evaluated. Data collected will come from synchronous and asynchronous forms of interaction plus teacher interviews on the impact of the team on the students’ classroom interactions. Problem solving will be analyzed in terms of the group’s process of mathematical thinking and interaction. The qualitative data will also be analyzed looking at the impact of group heterogeneity on individual learning and the effect of group composition on collaborative learning moments. The better performing groups of each year will be compared with the method of group selection. Quantitative data will be used to determine the extent to which involvement in collaborative small teams working on PoWs led to a general increase in usage of the MFDL and participation in the MFDL community. In year II, creator team and design team evaluation will follow the pattern set up in year I. Two creator and two design teams will be studied each year. Data will be collected through synchronous and asynchronous communication and interviews with group members. Data analysis will follow the pattern in year I and will be done to identify effective teams as well as teams that enhance the development of individual members of the group.

5. Anticipated Results & Impact

5.1. Dissemination & Outcomes

Dissemination of Project results, both in the US and in Europe, is built into the Project design. Dissemination to the international research community, to practicing educators and to the public generally will take place primarily through the following mechanisms:

- Involvement of international researchers. Approximately two dozen researchers will be intimately involved in this Project, primarily through the design teams. Many of their graduate students will also be involved.

- Workshops at international conferences. The Project will sponsor at least one workshop to bring together international and American
researchers in the design teams. This may be coordinated with international conferences on education such as CSCL, ICLS, AERA and EARLI. Most of the researchers involved in this Project regularly attend these conferences and present at them. These conferences will be primary sites for the presentation of results from this Project. Project staff will submit papers and organize presentations about the Project results at these conferences.

- Involvement of teachers and student teachers. Perhaps two dozen teachers and student teachers will be intimately involved in this Project, primarily through the creator teams. As the results of this Project become part of MFDL’s regular services, increasing numbers of teachers and student teachers will participate in spontaneously formed curricular workgroups.

- The MFDL virtual community. This is a rapidly growing community that already numbers over a million distinct individuals. They will learn about the results of this Project as collaborative problems become a regular feature of the MFDL and as community participants are automatically invited to join small groups for collaborative learning of mathematics.

- The NSDL community. The MFCLE will be presented at NSDL gatherings and through NSDL communications as a model for collaborative services in digital libraries.

5.2. Sustainability & Contribution

The results of this Project, particularly the MFCLE service, will be fully incorporated in the MFDL. The MFDL is a permanent program within Drexel University, so that services developed in this Project will continue to exist and be used indefinitely. Although the MFDL receives grants to engage in research and service expansion, it strives to develop revenue sources to sustain existing services. The collaboration services of this Project will contribute to building new lines of revenue, including contracted services with school districts for which MFDL will provide custom collaboration services and support.

5.3. Integration of Research & Education

The MFDL itself integrates research and education. It provides resources and services to support math education over a broad range of school grades, as well as meeting educational needs of employees, mathematicians and lifelong learners. The MFDL organization is heavily involved in research on
digital libraries, often in conjunction with academics at Drexel University (like the co-PIs in this Project).

The specific content of this Project applies technologies at the forefront of CSCL and CSCW research to educational needs. The emphasis on small group collaboration as an important mode of educational practice also comes out of recent research in learning theory.

5.4. Integrating Diversity

A central Project hypothesis is that groups integrating specific kinds of diversity learn better. The MFCLE software will be designed to optimize diversity during the group formation process.

5.5. Intellectual Merit

This Project creatively combines leading-edge collaboration technologies with one of the most popular services of a successful digital library to provide a model of support services for collaborative digital library usage. The Project brings together four co-PIs with the required mix of expertise, along with teams of engaged educators and international researchers.

This Project systematically explores an important open challenge of the Internet: how to foster effective collaborative online learning in digital libraries. It joins the multidisciplinary expertise of the international CSCL community with the practical success of the MFDL to study how to mediate the growth of a large virtual learning community, and to design, develop and assess tools for the online support of small workgroups acquiring, managing and negotiating knowledge.

5.6. Broader Impacts

The Project develops collaboration services for digital libraries, providing a sustainable model. It promotes the involvement of geographically isolated, disadvantaged and disabled students, distributed teachers and international researchers by inviting them into collaborative learning teams hosted, supported and informed by a digital library. The MFDL PoW service already attracts hundreds of thousands of people to the digital library and its resources; with the MFCLE support, more people will become more intensely involved in the user community. Other digital libraries can copy this model, providing services that attract visitors to specific resources and involve them in group activities. This Project pioneers a path for enhancing NSDL impact, building effective virtual learning communities.
The MFCLE software, with automated formation of small groups and with support for interactions that develop deep understanding of mathematics, will be suggestive for virtual learning communities in other domains and other digital libraries. This model provides opportunities for students, teachers and researchers excluded from collaborative learning due to geographic isolation, disadvantaged schools, physical disability, discrimination and other physical or social factors. The model stimulates both student motivation and teacher development, transforming interest in mathematics from a potential social stigma into a bridge to global friendships.

References


Proposals for Research

mathematical problem solving (pp. 32-60). Hillsdale, NJ: Lawrence Erlbaum Associates.


Part II: Other Proposals at Drexel University
DR K-12: Computer-Supported Math Discourse Among Teachers and Students

This full research-and-development project designs, develops and tests an interrelated system of technological, pedagogical and analytic components to provide a range of opportunities for middle- and high-school students to engage in significant mathematical discourse (DR K-12 challenge 2); it catalyzes and supports these opportunities by enhancing the ability of in-service teachers to engage in, appreciate and foster math-problem-exploration and math-discourse skills in their students (DR K-12 challenge 3). The project addresses the core STEM discipline of mathematics by motivating the identification, comprehension and enjoyment of mathematical discourse skills through socially interactive, collaborative learning experiences involving pedagogically organized series of stimulating, skill-appropriate problems using computer-based visualization/exploration and small-group math-problem discussion.

The project’s design-based-research approach crafts a socio-technical educational model to provide a comprehensive, practical package of tools and techniques for classroom teachers and students, which integrates and refines a number of mutually supportive components: (a) **Innovative technology**: A custom, open-source virtual learning environment that integrates synchronous and asynchronous media with the first multi-user dynamic-math-visualization application. (b) **Curricular resources**: Problem-based learning topics in specific areas of mathematics designed to help teachers tune rich math problems to local texts or curriculum and to guide student exploration. (c) **In-service teacher professional development**: Practicing teachers in online masters programs are mentored to understand and model the innovative technologies and pedagogies by doing collaborative problem posing/exploring/solving and engaging in collaborative reflection on the math discourse in their logged interactions. (d) **Middle- and high-school students**: The teachers introduce the model, technology and resources into their classrooms.

The project builds on and integrates previous work of the PIs, including: the discourse-analysis-based theory of group cognition (Stahl, 2006); the Virtual Math Teams learning environment developed, analyzed and evaluated in
(Stahl, 2009b); curricular materials and dynamic math visualization software of GeoGebra, adapted to flexible multi-user collaborative learning; online professional development and online mentoring of in-service math teachers at the Math Forum and at the Drexel and Rutgers-Newark schools of education; and the adaptation of conversation analysis to text-based chat interaction analysis, designed to highlight how collaborative problem solving or group knowledge building takes place. The project adapts components that have been explored, prototyped, or piloted by the PIs to classroom use. Project key personnel and Advisory Committee members bring expertise and experience in educational software R&D; math problem-set adaptation, dissemination and mentoring; in-service math teacher training; online math resources, collaborative learning, problem-based learning and dynamic math; design-based educational research management and evaluation; theory of knowledge building in small groups and in online communities. They also bring opportunities for national deployment and scaling up.

**Intellectual merit.** This project integrates leading-edge cyber-learning-environment technology incorporating innovative collaborative math exploration tools with educational approaches based on current directions in the learning sciences. It approaches this through a systematic iterative process of co-evolving the technology and curricular resources in the context of engaging, reflective collaborative-learning experiences of significant mathematical discourse by in-service teachers and their students. It thereby advances theory, technology and practice within real-world educational settings to forge a coherent research-based approach to math education appropriate to today’s challenges and potentials.

**Broader impact.** The project designs, tests, integrates, evaluates and disseminates technology, curricular resources, pedagogical methods and analytic tools for use in math-teacher professional-development programs, classrooms of math students, home-schooling networks, online schools and the Math Forum community (over three million visits per month). Project results will support the use of math exploration technology within collaborative math-discourse approaches at diverse schools nationally through their spread to in-service teacher-training programs and services—bringing practical cyber-learning of math to at-risk and isolated math students. It documents the potential impact on both teachers and students of this computer-supported math-discourse approach quantitatively and qualitatively.
Project Description

Mathematics education in the future faces enormous opportunities from the availability of ubiquitous digital networks, from innovative educational approaches based on theories of collaborative learning and from rich resources for interactive, online, dynamic math exploration. The fact that more and more teachers and students are learning online—with distance education, online masters programs, home schooling, online high schools, etc.—makes the incorporation of virtual collaborative learning environments a natural trend. A major issue in realizing these opportunities on a broad scale in schools is empowering teachers to appreciate and engage in the new approaches, and supporting them with appropriate tools, models and resources for practical instructional usage.

This project therefore proposes to develop a model of professional development and a suite of supports for math teachers. It will design, test, evaluate and refine a virtual learning environment that integrates synchronous and asynchronous media with an innovative multi-user version of a dynamic math visualization and exploration toolbox. Online teams of in-service teachers will be introduced to the collaborative exploration of Common Core State Standards-based math topics in this environment. They will then be guided in reflection on their own team’s discourse with the use of chat-replaying tools. As they become familiar with the use of the technology and with the nature of collaborative math discourse, some of the trained teachers will mentor other teachers through a similar process of engagement. Also, they will introduce their students—primarily in diverse urban schools—to experiences of mathematical exploration and to reflection on math-team discourse. The model of math teacher professional development and of student collaborative math learning centers on the production of significant math discourse.

Theoretical Framework: Math Cognition as Math Discourse

To mathematicians since Euclid, math represents the paradigm of creative intellectual activity. Its methods set the standard throughout Western civilization for rigorous thought, problem solving and argumentation. Many of us teach math in part to instill in students a sense of deductive reasoning. Yet, too many students—and even some math teachers—end up saying that they “hate math” and that “math is boring” or that they are “not good at math” (Boaler, 2008; Lockhart, 2009). They have somehow missed the
intellectual math experience—and this may limit their lifelong interest in science, engineering and technology. According to a recent “cognitive history” of the origin of deduction in Greek mathematics (Netz, 1999), the primordial math experience in 5th and 4th Century BC was based on the confluence of labeled geometric diagrams (shared visualizations) and a language of written mathematics (asynchronous collaborative discourse), which supported the rapid evolution of math cognition in a small community of math discourse around the Mediterranean, profoundly extending mathematics and Western thinking.

The vision behind our project is to foster communities of math discourse in networks of math teachers, in classrooms of K-12 math students and in online communities associated with the Math Forum. We want to leverage the potential of networked computers and dynamic math applications to catalyze groups of people exploring math and experiencing the intellectual excitement that Euclid’s colleagues felt—refining and testing emerging 21st Century media of collaborative math discourse and shared math visualization to support math discourse in both formal and informal settings and groupings. Those members of the project team who teach math teachers masters-level courses and professional-development workshops—and others—have found that many people teaching K-12 math have had little experience themselves participating in processes of mathematical exploration and discovery (Krause, 1986; Livingston, 1999; Silverman & Thompson, 2008). This project is designed to provide teachers with firsthand experiences and to mentor them in guiding their students to engage in rich math discourses that go beyond generating numeric answers to supply math reasoning and to draw conceptual connections (Briedenbach et al., 1992; Carlson, 1998; Carlson et al., 2002; Monk, 1992; Thompson, 1994).

The learning sciences have transformed our vision of education in the future (Sawyer, 2006; Stahl, Koschmann & Suthers, 2006). New theories of mathematical cognition (Bransford, Brown & Cocking, 1999; Brown & Campione, 1994; Greeno & Goldman, 1998; Hall & Stevens, 1995; Lakatos, 1976; Lemke, 1993; Livingston, 1999) and math education (Boaler, 2008; Cobb, Yackel & McClain, 2000; Lockhart, 2009; Moss & Beatty, 2006), in particular, stress collaborative knowledge building (Bereiter, 2002; Scardamalia & Bereiter, 1996; Schwarz, 1997), problem-based learning (Barrows, 1994; Koschmann, Glenn & Conlee, 1997), dialogicality (Wegerif, 2007), argumentation (Andriessen, Baker & Suthers, 2003), accountable talk (Michaels, O’Connor & Resnick, 2008), group cognition (Stahl, 2006) and engagement in math discourse (Sfard, 2008; Stahl, 2008). These approaches place the focus on problem solving, problem posing, exploration of alternative strategies, inter-animation of perspectives, verbal
articulation, argumentation, deductive reasoning and heuristics as features of **significant math discourse** (Maher, Powell & Uptegrove, 2010; Powell, Francisco & Maher, 2003; Powell & López, 1989).

To learn math is to participate in a mathematical discourse community (Lave & Wenger, 1991; Sfard, 2008; Vygotsky, 1930/1978) that includes people literate in and conversant with topics in mathematics beyond basic arithmetic. Learning to “speak math” is best done by sharing and discussing rich math experiences within a supportive math discourse community (Papert, 1980; van Aalst, 2009). By articulating thinking and learning in text, students make their cognition public and visible. This calls for a reorientation of the teaching profession to facilitate dialogical student practices as well as requiring content and resources to guide and support the student discourses. Teachers and students must learn to adopt, appreciate and take advantage of the visible nature of collaborative learning. The emphasis on text-based collaborative learning can be well supported by computers with appropriate computer-supported collaborative learning (CSCL) software, such as that prototyped in the Virtual Math Teams (VMT) Project (Stahl, 2009b).

**Research Project Goal, Hypothesis and Components**

**Project Goal**

To incrementally refine a research-based, classroom-tested model of computer-supported, resource-supported math education through **shared visualizations** and **collaborative discourse** by groups of mentored **teachers** and groups of their **students**—by designing, developing and testing: (i) a **discourse-based model** of math-teacher professional development and mentoring support; (ii) customized **technology** for computer support of **shared math visualization** and **joint exploration**; and (iii) adaptable, standards-based math-content teaching **resources** for middle-school and high-school students, guidelines for group **collaboration** and **accountable talk**, tools for **reflection** on discourse and networks of on-going **mentoring** relationships for math teachers.
Research Hypothesis

The project is based on an hypothesis, which it will test, concerning how to increase the quality and quantity of significant math discourse among math teachers and K-12 students:

Indicators of math learning (by groups of teachers and groups of their students)—such as group discussion of math content, problem posing/exploring/solving, explanation of math moves, visualization or investigation of multiple representations, and reflexive analysis of group math work—can be increased through (i) a math-discourse-based model of in-service teacher professional development supported by and integrated with use of (ii) a multi-user version of dynamic mathematics technology integrated in a rich online learning environment to support shared visualization and joint exploration of mathematical topics and (iii) mentoring relationships, collaboration and accountable talk guidelines, and curricular resources for online professional-development courses, K-12 classes and formal and informal online math communities.

This hypothesis is intended to guide iterative cycles of trial and analysis in design-based research (design, develop and test—not to prove efficacy and effectiveness). It will assess the effect of the combination of project components—because in such a socio-technical system the effect of introducing the technology is highly dependent upon the mentoring and the use of appropriate resources.

The hypothesis centers on measurements of group math discourse rather than on assessment of individual learning of math content—in accordance with the socio-cultural theory that effective individual math learning can be an indirect product of participation in group math discourse (Lave & Wenger, 1991; Sfard, 1998; 2008; Stahl, 2006; Vygotsky, 1930/1978). Vygotsky's notion of the zone of proximal development suggests that students may be able to engage in mathematical work within groups at a level that they will not be able to engage in for a couple years as individuals—and that such group work can be essential for the individual development in the long run (Vygotsky, 1930/1978, pp. 84-91). As a result, there is a need to assess the educational effectiveness of group interactions as such, beyond pre/post tests of the individuals. In addition, the striking finding within CSCL research of productive failure (Barron, 2003; Kapur & Kinzer, 2009; Patak et al., 2011; Schwartz, 1995) shows that there can be a paradoxical inverse relationship between measures of successful learning by small groups versus by the individual members of those groups because of
group processes that reveal deep mathematical relationships but that do not lead immediately to high test scores of the individuals. For these reasons, the project evaluates its goal in terms of the quantity and quality of the math discourse that takes place during the small-group problem-solving interactions, looking for hypothesized increases for groups as they participate and in successive project years as the model, technology and resources are iteratively developed.

(i) Model of Math Education

The proposed project will design, develop and test a model of math education through collaborative math problem proposing/exploring/solving, by involving in-service teachers in first-hand mathematical experiences and helping them to reflect on their own learning experiences. Then they will try out the model with their students, while receiving mentoring and support from the project. The collaborative model of math education stresses math discourse. In this project, groups of teachers and groups of students will do math problem solving collaboratively and then reflect on the logs of their discourse to identify key moves. We propose using teachers’ and their students’ original mathematical conversations as “didactic objects” (Thompson, 2002) designed to support “decentering” (Wolvin & Coakley, 1993) and “collective reflection” (Cobb et al., 1997) on particular aspects of their math discussion. The discourse-centered model of math education will structure learners in small teams and will provide mentoring to guide the team’s mathematical exploration, discourse and learning. Math Forum staff and other project team members will provide initial mentoring to the first cohorts of teachers, who will in turn mentor subsequent cohorts of teachers as well as students in their own classes. A permanent support network will be established to provide sustainability of project accomplishments. The teachers who are trained in this project will be encouraged—initially by paying them—to participate in teacher networks, including national and international networks of teachers, supporting broadening dissemination of the discourse model of math education.

(ii) Online Math Collaborative Learning Environment

The proposed project will design, develop and test two forms of technology to support math learning with collaborative and interactive tools for cyberlearning: (a) computer-supported collaborative learning (CSCL) software and (b) dynamic mathematics (software that allows users to manipulate geometric diagrams, equations, etc.). (a) CSCL provides virtual learning environments in which teams of students can interact
synchronously and asynchronously to build knowledge together. This student-centered approach has many advantages, including increased motivation, sharing of skills, engaging in significant discourse and practicing teamwork. This project will adapt and extend the Virtual Math Teams (VMT) environment already prototyped and tested by the PIs (Stahl, 2009b). (b) Dynamic math (such as Geometer’s Sketchpad, Mathematica, Cabri or GeoGebra) has already profoundly impacted math education (Goldenberg, 1995; Hoyles & Noss, 1994; King & Schattschneider, 1997; Laborde, 1998; Myers, 2009; Scher, 2002), with Geometer’s Sketchpad and GeoGebra used in many US classrooms and globally. Yet, research on math education has not analyzed how students use dynamic math tools in sufficient detail (compare Çakır, Zemel & Stahl, 2009; Stahl, 2009b). GeoGebra (http://www.geogebra.org) is an open-source system for dynamic geometry, algebra and beginning calculus—including trigonometry, conics, matrices, graphing and Euclidean constructions. It offers multiple representations of objects in its graphics, algebra and spreadsheet views that are all dynamically linked, making GeoGebra a particularly flexible tool for exploration. Working with the developers of GeoGebra, this project will provide the first multi-user version of dynamic math, so that teacher teams and student teams can explore math collaboratively; it will integrate this into the larger VMT virtual collaborative-learning environment with text chat and wiki to support persistent discourses about math—that can be shared, reflected on and researched. (For a demo of the prototype system, go to http://vmt.mathforum.org/VMTLobby. Log in as “guest” with password “guest”. The Lobby should open showing the List of All Rooms. Select Project “VMT Research”. Click on "Apply filters". Open “Geometry”. Open “Polygons". Click on "GeoGebra Demo Room" Eventually a JavaWebStart chat room should open. Explore its different tabs and functions.)
Figure 1. A demo (not real student interaction data) GeoGebra construction created and discussed collaboratively in a proof-of-concept multi-user prototype of the project’s learning environment, based on the VMT system. The VMT system includes (not shown here): a Lobby with social networking and tools for teachers, integration with a wiki, and Web browsers.

(iii) Curricular Resources

The proposed project will design, develop and test resources to support teachers and students in their interactive explorations of rich math problems (e.g., open-ended problems with multiple possible solution approaches and many potential extensions to explore). Three kinds of resources are: (1) Curriculum packages in domains of K-12 math, building on existing NSF-funded and community-based sources (see Error! Hyperlink reference not valid., Error! Hyperlink reference not valid., and Error! Hyperlink reference not valid.). The curriculum will be based largely on classroom-tested problems using dynamic-math software and integrated with popular math textbooks (e.g., Everyday Mathematics, Investigations in Number, Data and Space, Mathematics in Context, Connected Mathematics, Interactive Mathematics Program, Core-Plus Mathematics, Simms Integrated Mathematics and textbooks from McDougal Littell or Glencoe), but adjusted by experienced Math Forum staff for collaborative online usage. It will be aligned with the recommendations of the Common Core State Standards for Mathematics and the new NCTM volumes, Focus in High School Mathematics: Reasoning and Sense Making in
Algebra/Geometry. Teachers will be mentored in adapting the content of their local curriculum to collaborative online student exploration, whether using GeoGebra or not. (2) Guidelines, suggestions and examples for collaborative learning, knowledge building and math exploration will be published. This will feature “accountable talk” guidelines for math discourse. (3) Training resources in understanding online math discourse will be developed to help teachers and students identify examples of productive inquiry moves, etc., to foster reflection on logs of their math discourses. These broad categories of resources will encapsulate the expertise of the project team in problem design, collaboration mentoring and discourse analysis, producing documents that can be used by a gradually growing community of math teachers and students. The content of these resources will build on experience at the Math Forum, the VMT Project, the teacher professional-development programs at Drexel and Rutgers and the related research literature. The content will be elaborated, tested, evaluated and refined—and then published as project deliverables.

Results from Prior NSF Support

The proposed project grows out of the successful Virtual Math Teams (VMT) Project. This is a several-year NSF project (awards DUE-0333493, IERI-0325447, SBE-0518477, DRL-0723580) that developed an open-source virtual learning environment for math students. The system integrated a social-networking portal, synchronous text chat, a shared whiteboard, an asynchronous wiki, a referencing tool, mathML expressions and a web browser. Student actions and chat postings are automatically logged to be replayed for analysis. Over a thousand student-hours of piloted usage were logged. A qualitative micro-analytic approach to interaction analysis was developed based on ethnomethodologically inspired conversation analysis (Garfinkel, 1967; Sacks, 1962/1995; Stahl, 2009a; 2009c; Zemel, Çakir & Stahl, 2009). A large number of publications have appeared from the project (see http://GerryStahl.net/vmt/pubs.html), including 2 books (Stahl, 2006; 2009b) and 6 doctoral dissertations (Çakir, 2009; Litz, 2007; Mühlpfordt, 2008; Sarmiento-Klapper, 2009; Wee, 2009; Zhou, 2010).

The VMT Project pioneered the study of online collaborative math discourse—both its nature and modes of computer support for it. The 28 studies in (Stahl, 2009b) present some of the most important of the 169 publications related to the project. They include a number of dissertation-level case studies of interactions in the VMT environment by middle-school,
high-school and junior-college students, which analyze: how math problem solving can be effectively conducted collaboratively among students who have never met face-to-face; how the structure of text chat interaction differs from spoken conversation; how the media of graphical diagrams, textual narratives and symbolic representations can be intimately interwoven to build deep math understanding; how deictic referencing is important to establishing shared understanding; how students co-construct a joint problem space; how collaborative meaning making and knowledge building are accomplished in detail; how online math discourse can be supported by a software environment that integrates synchronous and asynchronous media with specialized math tools; and how a methodology based on interaction analysis can be used for a science of group cognition.

The VMT Project was structured as design-based research, with the technology, research and theory co-evolving through dozens of iterations. The VMT Project demonstrated both the practicality of the proposed project and the need for it. While the VMT Project prototyped a rich cyber-learning environment and studied student interaction, it did not develop the range of supports that we know are needed for classroom use: robust software, problem sets, guidelines, etc. Furthermore, it did not include a dynamic-math component. The VMT Project provides a solid starting point for the proposed project and documents the need for further technological development, enhanced support for dynamic math, curricular models and training of in-service teachers. The design, development and testing of these logical next steps are needed to enable a powerful and innovative form of math education to be offered in a practical form to K-12 schools through education schools and to the public through the Math Forum.

Prior NSF support of the Math Forum has developed a successful approach to online mentoring of math teachers and their students. Since 1993, the Math Forum has mentored over 100,000 students, conducting hundreds of workshops, summer institutes and school-improvement contracts. Recently, it has successfully completed the Virtual Fieldwork, Online Mentoring, and Teacher Workshop Model projects (NSF DUE-0717732, DUE-0127516 and DUE-0532796). Mixed-methods studies of these have shown the surprising result that the online mentoring of K-12 pre- and in-service teachers had a more positive effect for teachers with low math self-efficacy (Renninger et al., in press). This is due to the non-linear and flexible format of online discussion—suggesting that online collaboration may well help at-risk math students at least as much as those with higher math self-efficacy. Math Forum approaches are making inroads with a population of people who most would think will not change (Renninger et al., 2010). In the proposed project, Math Forum workshops for teachers will complement and feed
teachers into the courses at Rutgers and Drexel. The workshops will also train mentors and seed the on-going teacher network.

**Research and Development Design**

The proposed project adopts an iterative design-based-research approach to design, develop and test innovative curriculum materials, technologies, teaching methods and models for teacher in-service professional development and K-12 student instruction. The project develops a socio-technical educational model that evolves and integrates a number of mutually supportive components, each of which has previously been explored in a preliminary way by one or more of the PIs. However, the components have not previously been integrated into a scalable model of math education. The proposed project brings together the PIs, other necessary senior staff and advisors with the resources to begin to systematically test, refine, validate and disseminate the integrated model. There are several areas of work:

(a) **A model of math education as computer-supported math discourse.** The model incorporates: (b) innovative technology for collaborative math discourse, (c) support for shared mathematical visualizations and (d) curricular materials to stimulate and guide math discourse. The model includes three successive project targets: (e) in-service teacher professional development, (f) middle- and high-school math education and (g) broader virtual math-discourse communities.
Figures 2 and 3. Images of actual student online collaborative work on patterns. In Figure 2, a student points from a chat message to a smallest hexagon pattern composed of 6 triangles illustrating VMT's unique integration of chat and whiteboard with its deictic reference tool. Figure 3 shows the Replayer tool interface across the bottom.

(b) Innovative technology for collaborative math discourse. The VMT Project developed a research prototype of a custom, open-source virtual learning environment that integrates synchronous (text chat, shared whiteboard, dynamic math exploration, shared web-browser) and asynchronous (a community wiki, a social-networking portal) media to support math visualization and collaborative discourse by virtual math teams. This prototype was adequate for extensive testing in multiple iterations, as well as limited use by select teachers in their classrooms as part of research trials. As part of the proposed project, we will implement, test and refine new interfaces for teachers, mentors and administrators. These will allow teachers to register a number of students at once, set up multiple copies of interaction rooms for multiple small groups of students, monitor activity in rooms, respond to problem behavior online and review reports of student activity. New functionality will also make it easier for students to document their online work (e.g., in the project wiki or in Word documents, Excel spreadsheets and PowerPoint slides) with log excerpts and images of constructions. Support for researchers will facilitate researchers in the project as well as colleagues outside the project to easily replay sessions of student interaction.
The VMT Project was widely recognized as an important example of synchronous support for online collaboration and was studied by several international researchers (Error! Hyperlink reference not valid.); it is expected that the proposed project will be of even more interest, particularly within the math education research community. The VMT Replayer allows complete replay of a user session, including all actions and system notices, as though the session was digitally video-recorded. The researcher’s view is guaranteed to be identical to the user’s view since it is generated from the same data as sent to a client computer. The log information will be made available in convenient textual formats for student reflection and reporting as well as for researcher analysis. New functionality to be explored includes automated feedback agents and displays, increased integration so math objects can be moved easily from the synchronous tabs (chat, whiteboard, summary, GeoGebra, web browsers) to asynchronous components (wiki pages, email, documents), as well as refinement of the interface. The system will be released as open source on SourceForge so that others can deploy it on their own servers or extend the software to meet their own educational needs. The Math Forum will maintain the system as a permanent service, so that users can easily create topics for chat rooms and invite other users to collaborate.

(c) Support for shared mathematical visualizations. The project will port GeoGebra—a comprehensive and well established application for dynamic-math exploration—to the virtual learning environment described above. It will make the application fully multi-user. It will integrate the application in a tab of the environment (see Figure 1 above). As previously described, GeoGebra is a particularly appropriate dynamic-math application for this project because its source code is freely available as open source, there is an active international development community to support on-going development, the lead developer and the founder are committed to consult on this project, the application supports a wide range of math from algebra and geometry construction to calculus and 3-D, GeoGebra has won international prizes, it has been translated into about 50 languages and it has received on-going NSF support. Like all other dynamic-math applications, GeoGebra currently exists only as a single-user application. While users can send their static constructions to each other, display screen images, or awkwardly include a view of the GeoGebra application within other environments through screen sharing (e.g., in Blackboard, Moodle, Elluminate, etc.), only one person can dynamically manipulate the construction. Our port converted GeoGebra to a client-server architecture, allowing multiple distributed users to manipulate constructions and to all observe everyone’s actions in real time. Every action in the GeoGebra tab
will be immediately broadcast by the server to all collaborating clients (and logged in detail for replay and research). We have been exploring turn-taking mechanisms (see Figure 4) to avoid conflicts in the construction and modification of GeoGebra drawings; although it is important in synchronous chat to allow multiple users to type simultaneously, we have found that it is natural for a group to allow one member at a time to change a graphical construction and for group members to take turns editing and rearranging.

Figure 4. The GeoGebra tab with turn-taking button to avoid conflicts.

Incorporation of GeoGebra in the VMT environment framework allows users to engage in text chat while manipulating the construction. Importantly, users can graphically point from a chat posting to an area of the construction that they want to index (see Figure 2)—an important support for math discourse that is unique to VMT. They can also scroll back and forth through the history of the GeoGebra construction, animating its evolution—a powerful way to explore many mathematical relationships. In addition, a complete record of the collaborative construction is available to the participants, their teachers and project researchers, allowing them all to analyze and reflect upon the complete interaction, including the construction actions synchronized with the chat. We have already completed a prototype port of GeoGebra to VMT in order to confirm its feasibility. It provides an exciting collaborative experience. The port now needs to be engineered in a robust way, incorporating all of the GeoGebra functionality (including import and export compatible with standard GeoGebra and Geometer’s Sketchpad to facilitate sharing of constructions, and a full menu system to support learning by new users). In Year II of the project, we will incorporate the extended GeoGebra 4.0 functionality that will be released by then, including support for inequalities and CAS (computer algebra system like
Mathematica, Maple, or the TI-Nspire CAS calculator). The project will produce a refined and tested multi-user version of GeoGebra and will release it as open source.

(d) **Curricular materials to stimulate and guide math discourse.** Problem-based learning (PBL) materials in areas of mathematics like algebra, combinatorics and geometry will be adapted from existing high quality curricula and piloted. These materials will define challenging math problems for collaborative online group exploration and help teachers to tune them to local student capabilities. The materials will allow students to explore rich but accessible problems taken from topic domains covered in their textbooks and in the Common Core State Standards. The PBL approach involves mentors who are trained to guide student exploration and to steer collaborative student groups to address their joint learning issues (Barrows, 1994; Hmelo-Silver, 2004; Hmelo-Silver & Barrows, 2008; Koschmann, Glenn & Conlee, 2000). Project team members and others have developed some model math problems (Krause, 1986; Math Forum & Wolk-Stanley, 2003a; 2003b; 2004a; 2004b; Powell, Lai & O’Hara, 2009). The Math Forum has years of Problems-of-the-Week in several areas of school mathematics, which can be adapted to online group collaboration. Much curriculum has been developed with NSF funding for dynamic-math applications like GeoGebra and Geometer’s Sketchpad, including lessons tied to state standards and intended to support popular textbooks through student hands-on exploration. The project will facilitate classroom teacher use of such resources in this new learning context. The team has already prototyped a series of problems that consecutively explore issues of combinatorics; along with the problems, a teachers’ guide contains concrete suggestions on how to adapt the problems for different kinds of student teams (Powell, Lai & O’Hara, 2009). The problems in this document were tested in the VMT Project and in high-school classrooms of teachers studying at Rutgers. Sets of problems correlated to textbooks and to the Common Core State Standards will be compiled, some taking advantage of GeoGebra. Additional resources will be developed to train teachers and students in mentoring techniques, in collaboration skills and in math-discourse skills. All these resources will be tested and produced in publically available online documents as project deliverables. These and other math problems will be incorporated in the VMT Lobby's library of Topics, to be available to students in home-schooling and informal-learning situations.

(e) **In-service teacher professional development.** To effectively change education in schools, teachers must be prepared to understand and to learn how to model use of the innovative technologies and pedagogies. Practicing
teachers rarely find time to engage in learning processes capable of transforming their teaching practice and they seldom are able to introduce major new approaches in their highly constrained curricula. This project therefore involves in-service teachers when they have scheduled time to pursue masters-level professional-development courses. It starts by involving them during their regular courses (taken online) in online collaborative problem solving using the project’s software technology and curricular approach—(a), (b) and (c) above. Later course work involves them trying out what they have learned back in their own classrooms, within the context of their current curriculum; the project provides mentoring and resources to support this effort.

Both Drexel and Rutgers-Newark offer masters-level teacher-professional-development programs and courses in math education in online modes. The fact that these teachers will already be studying together online creates an ideal setting for the use of an online learning environment with dynamic-math support. These graduate programs have been designed, taught and directed by project co-PIs Silverman and Powell. The proposed project will allow these programs to develop, test and adopt the educational model of computer-supported math discourse. This model will be pioneered at these two schools of education, providing a collaborative interaction that will produce a more generalized result than would development at a single institution. It will also permit extended utilization of the online medium by, for instance, having teachers from both institutions working together on math topics in small groups and having them mentor teachers from each other’s institution. In the later years of the project, this model will be disseminated to other schools of education, partially through Advisory Committee members. The Math Forum has effectively implemented a similar model, incorporating its Online Mentoring Project modules into teacher education programs around the country.

The initial plan at Drexel University is to build on the existing MS in Mathematics Learning and Teaching (MS-MLT) program, which is already exclusively offered online. This program in math education was originally developed by co-PI Silverman and is taught primarily by him and Math Forum staff. For the first cohort of students under this project, Drexel will offer MTED775, “Special Topics: Supporting Math Learning through Computer-Supported Collaborative Discourse.” This course will be one required math-education elective for MS-MLT students and an elective for other professional-development students. Then two new education courses will be developed to make this model a part of the regular course offerings of the School of Education: MTED 651 (which will focus on teachers personally engaging in computer-supported, resource-supported
Proposals for Research

Collaborative discourse and reflection on both their activity and their learning) and MTED 652 (which will focus on supporting teachers to incorporate computer-supported, resource-supported collaborative discourse in their classes). MTED 652 will include resource development for teachers' classroom implementations. Each of these courses—which have been approved at Drexel pending funding of this project—will carry 3 quarter-credits.

The initial plan at Rutgers-Newark is to engage two cohorts each year of practicing teachers in a revised version of the online course in “Mathematics and Instructional Technology” taught by co-PI Powell. The goals of the course are three-fold: (1) to familiarize in-service teachers with the mathematical problem-solving and problem-posing activities of the online problem-exploration units in which their students will engage; (2) to deepen in-service teachers’ thinking about the effects of the collaborative environment on their own and their students’ thinking about mathematics (math objects, relations among objects and dynamics among relations), math reasoning and problem-solving heuristics; (3) to focus in-service teachers’ instructional attention on understanding and facilitating students’ discourse in mathematics. To accomplish these goals, the course will engage in-service teachers in a sequence of tasks, beginning with familiarizing them with the project online environment through involving them in mathematical activities using it, then engaging them in reviewing their session logs and finally having them plan how they will implement the model in their teaching.

(f) Middle- and high-school math education. The in-service teachers will introduce the technology and curricular resources that they used in their university classes into the classes they teach, often mixing students from different schools or cities in online teams to take advantage of being part of an online discourse community and to motivate the use of online media by students in face-to-face classrooms. The teachers will take the logs of their students’ interactions back to their professional-development sessions for on-going group analysis. They also will engage their students in reflection on their own logs, discussing how the math discourse surfaces mathematical insights and conceptual connections.

The curricular resources adapted by the project are designed to support classroom math activities by enhancing and reinforcing the core objectives covered in textbook readings and instructor-led activities. Resources include adaptation options and guidelines to help teachers tune problem sets to complement their core activities. For instance, the research-based textbooks, Mathematics in Context and Discovering Geometry, which are used in the
Philadelphia public school system, stress student investigation in order to construct conceptual understanding of key math concepts and the Common Core State Standards for Mathematics recommend that “students consider the available tools [such as] dynamic geometry software...to explore and deepen their understanding of concepts” (p.7). The project model builds on this approach, providing opportunities for students to explore and discuss topics online with peers from their own or other schools. The model provides: tools for dynamic, multi-user, graphical exploration; visual and numeric feedback on quantitative and qualitative changes during exploration; and a record of the exploration and accompanying discourse, which students can replay, reflect on and incorporate in reports—e.g., pasting log excerpts or screen images in their documents.

Reflection on interaction logs by teachers and students primarily involves trying to follow the problem-solving path of participants and to notice critical collaboration moves. They will be encouraged to look for examples of accountability to the group, to standards of math reasoning and to the characteristics of their math objects. They will look for instances where someone poses a productive inquiry that initiates effective group exploration—or where the group fails to come up with a useful proposal or fails to take up a proffered proposal. Examples will be culled and shared on the project wiki.

Although many project activities center on teacher professional development, the ultimate goal is to increase the quality and quantity of both teacher and student mathematical discourse. Therefore, teacher professional development will be oriented to improving the math discourse of their students. While the primary indicator of project success will be the identification of desirable mathematical discourse moves during problem solving by teachers and students, the project will also be concerned with changing student conceptions of math. It will survey a sample of teachers and students before and after their involvement in the project to compare self-reports of attitudes about math and about approaches to math instruction. In addition, some teachers and students will be asked as a final part of their course work to compose a brief reflection paper on their learning experience.

Most of the in-service teachers in the project come from the Philadelphia, Camden, Newark, New Brunswick and New York City areas. Thus, many of the classrooms that will be involved in the program are inner-city K-12 schools with high proportions of educationally at-risk and economically disadvantaged students; others are from near-by suburban and private
schools with contrasting student populations. The project educational model will therefore be tested in diverse, real-world settings.

Because teacher and student work on math problems will all take place in the online software environment, complete detailed logs will be available to the project staff, as well as to the students and teachers themselves. The logs can be reviewed and studied in detail with the Replayer software, as well as with various formats of log printouts. This will not only facilitate reflection by students and teachers on their own work, but also permit the documentation of interesting cases for teacher instruction and detailed analysis for project evaluation. The project will compile a portfolio of instructive case studies.

(g) **Broader math-discourse communities.** Once teachers studying at Drexel or Rutgers and their students become involved in online collaborative dynamic geometry and math discourse, teams will be set up that involve students from online schools, home-schooling networks or the Math Forum virtual community. This will yield data for generalizing project findings as well as stimulate the spontaneous generation of self-organizing communities of math discourse. This will primarily take place through contacts and presentations by project staff and the teachers who have been trained, as well as through the Math Forum and its large user community (3 million visits/month. The project technology and resources will be made publically available as an integral part of the Math Forum services in Years IV and V of the project. The VMT software environment is designed to support the viral spread of user communities across the Internet; the proposed project is intended to form a critical mass of users and topics to catalyze that process. The model of computer-supported math discourse will become institutionalized at Drexel and Rutgers, will be taken to other schools of education through Advisory Committee members and personal contacts of project staff, through Math Forum outreach, through the extensive active GeoGebra user community and through presentations at educational conferences and in related journals.

(h) **Group cognition theory.** When small groups engage in collaborative problem posing, exploring and solving, they can accomplish cognitive tasks interactively or transactively as a group. The project will analyze logs of student math work, shared visualizations and reflective discourse, using conversational analysis and statistical methods to study how students build on each other’s utterances, constructions and actions to accomplish mathematical cognition. Building on past work on group cognition (Çakır, Zemel & Stahl, 2009; Koschmann, Stahl & Zemel, 2009; Stahl, 2006; 2010a; 2010b), this will provide a contribution to theory of situated and
distributed cognition. In particular, analysis of the use of GeoGebra in a fully logged multi-user online environment with guidance in math discourse moves will pioneer in the development of theory of cognition in groups using dynamic-math tools, providing insight into math learning generally. Case studies and other findings with theoretical implications will be published.

### Project Phases, Milestones, Deliverables

<table>
<thead>
<tr>
<th>Model</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design, testing of courses for teachers</td>
<td>Integrate technology and resources fully in courses</td>
<td>Refine model based on formative evaluation</td>
<td>Publish articles about model</td>
<td>Disseminate model to other schools of education, etc.</td>
<td></td>
</tr>
<tr>
<td>Technology</td>
<td>Debug VMT; multi-user GeoGebra 3.3; menu system; implement full logging</td>
<td>New VMT functionality; multi-user GeoGebra 4.0; teacher admin &amp; monitoring supports</td>
<td>Release VMT as a Math Forum service; automate statistical analysis</td>
<td>Support VMT Open Source; develop feedback of analysis to participants</td>
<td>Disseminate VMT servers</td>
</tr>
<tr>
<td>Resources</td>
<td>Pilot teacher resources; develop student resources</td>
<td>Test teacher and student resources</td>
<td>Evaluate use of resources</td>
<td>Publish resources</td>
<td>Disseminate resources</td>
</tr>
<tr>
<td>Curricular materials</td>
<td>Review existing materials for GeoGebra and Geometer's Sketchpad</td>
<td>Compile problem sets aligned with standards and textbooks</td>
<td>Evaluate use of materials</td>
<td>Publish materials in formats for teachers, home schooling, distributed schooling</td>
<td>Disseminate materials</td>
</tr>
<tr>
<td>Teachers</td>
<td>Pilot model with 10 teachers in Drexel and Rutgers courses and 20 teachers in Math Forum workshops</td>
<td>Implement model with 35 teachers in Drexel and Rutgers courses and 40 teachers in Math Forum workshops</td>
<td>Evaluate model with 50 teachers in Drexel and Rutgers courses and 40 teachers in Math Forum workshops</td>
<td>Continue training with 60 teachers in Drexel and Rutgers courses and 40 teachers in Math Forum workshops</td>
<td>Evaluate teacher training in Drexel and Rutgers courses; continue training 40 teachers in Math Forum workshops</td>
</tr>
<tr>
<td>Students</td>
<td>Pilot with 25 students</td>
<td>Involve 750 students of teachers in courses and workshops; log series of sessions by student small groups</td>
<td>Involve 750 students of teachers in courses and workshops; log series of sessions by student small groups</td>
<td>Involve 750 students of teachers in courses and workshops; log series of sessions by student small groups</td>
<td>Evaluate changes in significant math discourse of student groups over time; within group and across cohorts</td>
</tr>
<tr>
<td>Mentoring</td>
<td>Prepare mentoring materials based on previous Math Forum mentoring projects</td>
<td>Pilot mentoring of teachers with 2 outstanding teachers</td>
<td>Increase to 5 teacher mentors</td>
<td>Increase to 10 teacher mentors</td>
<td>Increase to 15 teacher mentors</td>
</tr>
<tr>
<td>Theory</td>
<td>Validate coding scheme</td>
<td>Analyze discourse moves in logs</td>
<td>Conduct in-depth case studies and interviews</td>
<td>Compile best practices case studies</td>
<td>Develop theory of math group discourse</td>
</tr>
</tbody>
</table>

*Note: The table provides a detailed breakdown of the project phases, milestones, and deliverables, including specific tasks and timelines for each phase.*
Evaluation

Formative evaluation is a constant process built into the design of the project. As a design-based research effort, the over-riding research hypothesis listed at the start of this project description will be addressed by designing and exploring an iteratively refined solution—and by documenting its impact on the quantity and quality of math discourse by teachers and students. The interlocking components of the project will be reviewed at weekly project team meetings. Team meetings will include interaction analysis data sessions (Jordan & Henderson, 1995; Stahl, 2010a), in which the group collaboratively discusses new data from logs of teachers or students—and makes design decisions for refining the co-evolving components. The project team will discuss what seems to be working and what does not. It will decide what to modify for the next iteration. The project is complex, with many dependencies among its components and many shifting contextualities. A flexible approach like design-based research is needed to respond to a continuous formative evaluation and ongoing project modification.

The explicit evaluation effort will include semi-annual formative-assessment reports documenting: (a) project progress, (b) improvements in project outcomes and (c) plans for the next half year. The external Advisory Committee (AC) will review, discuss and respond to each report. The AC will meet annually to discuss project progress with the project team. The AC has expertise in mathematics education, research evaluation, teacher training, problem-based learning, conversation analysis, CSCL and virtual communities. Most AC members have been PIs on successful NSF grants in the learning sciences. The AC includes: Sharon Derry (Wisconsin), Cindy Hmelo-Silver (Rutgers-New Brunswick), Christopher Hoadley (NYU), Timothy Koschmann (Southern Illinois), Mary Marlino (UCAR), Kay McClain (Arizona State), K. Ann Renninger (Swarthmore), Lauren B. Resnick (LRDC, CMU), Carolyn Penstein Rosé (CMU), Anna Sfard (Haifa & Michigan State), Wesley Shumar (Drexel), Tamara Sumner (Colorado), Daniel D. Suthers (Hawaii), Alan Zemel (SUNY Albany). The external evaluator is Sukey Blanc (Senior Research Associate with Research for Action), who has led evaluations on projects such as the Metro Math MSP.

As discussed above, the research hypothesis focuses on the quantity and quality of math discourse at the group unit of analysis. Theories of the zone of proximal development, productive failure and group cognition argue that learning-related processes and phenomena at the group level may be different from those at the individual level. Other research has documented
the efficacy of dynamic-math visualization tools for *individual learning*; for instance, a study of geometry students in eleven Florida schools revealed a significant difference in the FCAT mathematics scores of students who were taught geometry using Geometer’s Sketchpad compared to those who used the traditional method—regardless of differences based on SES or gender (Myers, 2009). The proposed project has a different focus. The PI and colleagues have developed coding schemes and analysis approaches oriented to the *group unit of analysis* based on conversation analysis of adjacency pairs and longer sequences (Sacks, 1962/1995; Schegloff, 2007; Stahl, 2009b, Chs. 20, 22, 23, 26; 2011b; Stahl et al., 2011). This approach serves both quantitative and qualitative analysis, by simultaneously specifying the structure of meaningful discourse moves and providing countable categories of group interaction units, in order to document changes over time—comparing discourse characteristics in selected time slices within teams or across cohorts.

The project will automatically produce raw data in the form of log files of participant online interactions. The log files are anonymous, but allow tracking of individual users through consistent login handles. The VMT environment is instrumented to capture all user actions in the chat and whiteboard—this will be extended to multi-user GeoGebra. A database of all sessions is automatically maintained and provides spreadsheet logs in handy formats and Replayer files. Software tools will be used for automated and manual log analysis of discourse measures and their evolution during training. While low-level group processes (e.g., number, length and rate of chat postings and drawing actions in different time slices) can be tracked automatically and analyzed statistically, higher-level math-discourse processes have to be interpreted manually. The PI has on-going, NSF-supported collaborations with Carolyn Rosé of Carnegie-Mellon University’s intelligent tutoring group, exploring software agents in the VMT environment to provide student guidance and also investigating computer support for coding discourse moves in text chat, to aid and supplement manual analysis. Raw and coded logs will be maintained in a database to facilitate analysis of changes over time for groups across sessions and across successive cohorts of participants.

Quantitative analysis—based largely on the coding of discourse moves in teacher and student VMT logs—will track changes in key measures of significant math discourse. The project hypothesis will be operationalized as predicting an increase in specific measures as a given group works in the VMT environment during time slices across an academic term. Logs of the following groups involved in the project will be evaluated: (a) in-service teachers participating in Math Forum workshops, (b) teachers working
together as part of teacher professional development course work, (c) students guided by their teachers, (d) students working with other students as part of school classes and (e) students interacting with others informally at other schools or globally.

Discourse will be coded and measured along the following dimensions: (1) volume of discourse and level of participation, (2) percentage of on-task math discourse, (3) use of representations, (4) integration of chat and drawing, (5) use of accountable talk moves, (6) adoption of socio-mathematical norms and practices, (7) speaking meaningfully with explanation and argumentation, (8) involvement in posing, exploring and solving problems and (9) additional dimensions to be developed based on project experience. The theory of math learning through participation in math discourse (Sfard, 2008) specifies important mathematical discourse moves, such as encapsulation, reification, saming, routines, deeds, explorations and rituals. The theory of accountable talk (Michaels, O’Connor & Resnick, 2008; Resnick, 1988) specifies discourse moves that promote accountability to the group, to standards of math reasoning and to the characteristics of the math objects. Speaking meaningfully in math discourse “implies that responses are conceptually based, conclusions are supported by a mathematical argument and explanations include reference to the quantities in the problem context [as opposed to a focus on merely] describing the procedures and calculations used to determine the answer” (Clark, Moore & Carlson, 2008, p.298). Socio-mathematical norms include what counts as an acceptable, a justifiable, an easy, a clear, a different, an efficient, an elegant and a sophisticated explanation (Yackel, 1995; Yackel & Cobb, 1996). Mathematical practices emerge from interaction, are taken up by participants and are applied repeatedly (Medina, Suthers & Vatrapu, 2009; Stahl, 2011a). These dimensions of significant math discourse are associated with typical sentences and discourse moves that can be identified by coders. A coding scheme will be validated with acceptable inter-rater reliability, as in (Stahl, 2009b, Chs. 22, 23; 2011b).

Detailed interaction analysis of selected cases will show how the math discourse actually evolves. Quantitative analysis can establish the statistical significance of changes in learning outcomes, but it generally does not provide much insight into the mechanisms of change; these mechanisms will become visible in detailed case studies in which the specifics of the interactions can be studied. By combining quantitative and qualitative analysis of discourse transformations, the project evaluation will determine how the online interaction involves engagement in significant mathematical discourse. This will help researchers to determine what to try in subsequent cycles of research and will allow evaluators to judge project progress.
**Summative evaluation** will assess the degree to which the discourse of teams of teachers and of students reveals—through the quantity and quality of their math discourse—increased understanding and improved practice of mathematics. It will make sure that project products (software, mentoring guides, problem sets, masters courses, analysis tools, best practices case library and analyses of case studies from the data corpus) have been produced and made publicly available. It will assess the effectiveness of these products based on the analyses of their use by teachers and students as logged in the data corpus, using quantity and quality of the facilitated math discourse as a measure of success.

In addition to the quantitative and qualitative analysis of changes in significant mathematical discourse by groups of teachers and students involved in the project, there will be ethnographic observations of participants. The observations—including pre/post surveys, open-ended interviews and reflection reports—will be primarily conducted by co-PI Khoo and External Evaluator Blanc, both trained cultural anthropologists. The goal of these observations will be to establish—as much as possible from the user perspective of the project participants—the effectiveness of project interventions (the pedagogical model, the technology, the resources). Interviews with students and teachers will explore their changed attitudes toward mathematics and their insights into the nature of mathematical reasoning. This will be triangulated with the analysis of the math discourse of the same participants in specific time slices. Ethnographic observation of teachers will additionally explore to what extent they have come to feel that teaching math-discourse skills is key to fostering student math learning; to what extent they try to use the project model, technology and resources in their regular teaching; to what extent they intend to stay involved in support networks. The summative evaluation will report on these issues as well as the timely accomplishment of project tasks, training levels, dissemination efforts and project deliverables.

**Dissemination**

The primary avenues of dissemination will be: (a) through the Math Forum, (b) through Schools of Education, (c) through teacher professional associations, (d) through GeoGebra and dynamic math user communities and (e) through virtual learning communities, including home schooling and online schools.

(a) By the end of the project, the technology and the resources developed through the project will be publicly available as services of the Math Forum.
The Math Forum has been the premier online resource for mathematics teaching and learning for over 16 years. It has three million visits to the site each month; its digital library contains over a million web pages, mostly user generated (as a forerunner of the Web 2.0 philosophy). Public services (which typically started from NSF-supported research projects) have been made sustainable through support from Drexel University, fee-for-service programs and teacher training contracts. The Problem-of-the-Week (PoW) is the Math Forum’s core service and is subscribed to by many school districts. It is primarily oriented toward problem solving of challenging math problems by individual students. The result of the proposed project would be to extend this service with open-ended math problems for groups of students to explore collaboratively online. Teachers using the PoW service would be encouraged to involve their students in the new service, initially interacting with classmates, but eventually joining cross-school, national and international virtual math teams. Math Forum services typically support both formal and informal mathematics learning by teachers and students (Renninger & Farra, 2003; Renninger & Shumar, 2002b; 2004; Shumar & Renninger, 2002).

(b) Several of the co-PIs and Advisory Committee members (e.g., Powell, Silverman, Derry, Hmelo-Silver, Hoadley, Koschmann, McClain, Renninger and Sfard) teach at schools of education across the country—and are in contact with math educators at many more. The project accomplishments will influence the teacher professional-development programs in these centers. Teachers who are involved in the teacher professional-development components of this project will also spread project findings as early adopters at their graduate programs and K-12 schools. Ready access to project resources, models and technology at the Math Forum will facilitate general dissemination of innovative math education—including through the popular teacher discussion forums on the Math Forum website—to additional teacher professional development programs.

(c) The PIs and Math Forum are active in NCTM, AERA, PME, and PMENA and will present project findings at the annual conference for teachers of mathematics. Additionally, project researchers are prominent in the learning science communities around the ICLS, CSCL and other academic conferences and publish prolifically in academic and practitioner journals, books and conferences.

(d) Because it provides the first multi-user version of a dynamic-mathematics application, the project will be well known within the worldwide communities of GeoGebra and Geometer’s Sketchpad users. The project technology will all be available as open source, so that other
researchers and developers can build on it, modify it and install versions on their own servers. (The project technology is built on VMT and GeoGebra, both already available as open source at SourceForge.) Teachers, trainers and researchers who do not have the technical expertise to do this, can simply use the environment that is on the Math Forum servers; they can develop their own curriculum for it and can readily access detailed user logs from it. Features for administration of chat rooms will be built in to support local administration.

(e) For the sake of sustainability beyond the proposed project and to support further scale-up, it is important to establish an on-going network of teachers in the form of self-organizing communities (Renninger & Shumar, 2002a). As discussed above, this will begin with mentoring relationships between cohorts of teachers going through the project professional development. The mentoring relationship will grow into a mutual support network, in which teachers from the programs at both Drexel and Rutgers will share questions, case studies, best practices, curriculum, etc. Later in the project, this growing local network will connect with national and international teacher networks, such as Tapped-In (http://tappedin.org), the Knowledge Building Teacher Network (Chan, van Aalst & Law, 2009) and the Institute for Knowledge Innovation and Technology (Error! Hyperlink reference not valid.). These networks will disseminate use of the project services widely. We are aware of the issues in trying to build sustainable virtual learning communities (Barab, Kling & Gray, 2004) and will use an iterative approach. In addition, dissemination efforts will target organizations, consortia and networks of home schooling and of online schools.

**Expertise**

The proposed project brings together an interdisciplinary team of researchers, led by the PIs:

**PI, Gerry Stahl:** PI on the VMT Project. Author of *Group cognition: Computer support for building collaborative knowledge* and *Studying virtual math teams*. Founding editor of *International Journal of Computer-Supported Collaborative Learning*. He will have overall responsibility for the project.

**PI, Arthur Powell:** Chair of the Department of Urban Education at Rutgers-Newark and Associate Director of the Robert B. Davis Institute for Learning at Rutgers-New Brunswick. Specializes in problem solving, deductive reasoning and heuristics in math education. Expertise in analysis of learning
in digital video. Primary responsibility for teacher professional development at Rutgers.

**Co-PI, Jason Silverman:** Faculty member at the School of Education, Drexel University. Developed and teaches the online masters degree program in Mathematics Learning and Teaching at Drexel. Primary responsibility for teacher professional development at Drexel.

**Co-PI, Stephen Weimar:** Director of the Math Forum since 1994. Established track record as PI on multiple successful NSF grants. Responsible for integration with Math Forum services.

**Co-PI, Sean Goggins:** Brings a decade of collaborative and social software design and development team leadership. He will be primarily responsible for automated and statistical data analysis.

**Co-PI, Michael Khoo:** Trained in anthropology, he evaluated components of NSF NSDL digital libraries. He will coordinate the internal formative evaluation component of this project.

**Annie Fetter:** Co-founder of the Math Forum. Directs the Problem-of-the-Week. Has done professional development and written curriculum for the Geometer's Sketchpad software since it was created. She will be involved in training and mentoring the teachers and coordinating the classroom usage.

**Sukey Blanc:** Trained in urban anthropology, she studies mathematics and science education, educational equity and school reform. She is Senior Research Associate with Research in Action, a Philadelphia-based non-profit organization engaged in education research and evaluation, which since 1992 has worked with public school districts, educational institutions and community organizations to improve educational opportunities for those traditionally disadvantaged. She will work with the Advisory Committee and will be responsible for external formative and summative evaluation.

**The Math Forum.** This well established math education site, MathForum.org, has its office at Drexel University with program and technical staff to run services and to maintain the Internet technology. The staff has extensive experience in mentoring math teachers, training new mentors, designing math resources and supporting a huge user community. Most of the program staff are experienced classroom math teachers. The technical staff will be responsible for software development during the project and then for maintaining the project software during and beyond the lifetime of the project.

**The Advisory Committee.** The AC brings expertise in math education; educational psychology; quantitative analysis of learning outcomes,
motivation and attitudes; problem-based learning theory and analysis; CSCL; and online communities of learners. (See attached letters.)

References


ROLE: Studying Online Collaborative Learning at the Math Forum

Project Summary

The **project goal** is to determine through quantitative and qualitative methods if a small group of online learners can build knowledge that exceeds the knowledge of its individual members, and to study how this happens when it does.

The **project motivation** is that contemporary frameworks of learning theory stress the social (inter-personal, intersubjective) nature of knowledge building and recommend the fostering of collaborative learning in small groups. However, it is hard for many people to comprehend the idea of collaborative group knowledge distinct from and potentially exceeding the knowledge of the individual group members. There is little clear and incontrovertible evidence that online groups can construct knowledge that exceeds what any of the group members knows. It is important to study precisely how this can take place in order to design effective collaborative curriculum, technology and assessment methods. The proposed research does this in a concrete educational context.

The **project approach** is to develop an evidentiary base for computer-supported collaborative learning (CSCL) as applied in the Math Forum, a popular digital library used from classroom and home settings. By rigorously studying the learning process of online small groups solving specially designed problems in school mathematics, the project provides a strong research foundation for expansion of Math Forum services from an audience of primarily individual students to a virtual community of systematically matched groups of learners. It studies core social aspects of human learning (ROLE quadrant 2) within a context of math learning in educational settings (quadrant 3).

The **project objectives** are to study favorable conditions for building collaborative knowledge, including constructive criteria for group formation, for curriculum design and for technology support. The project will study the interaction of group and individual knowledge, and develop
rigorous, cumulative, reproducible and usable methods for analyzing and interpreting the relevant data. The central working hypothesis of this project is:

- H0 (collaborative learning hypothesis): A small online group of learners can (on occasion and under favorable conditions) build knowledge and understanding that exceeds that of its individual members. Other project working hypotheses guide specific objectives:
  - H1 (collaborative group hypothesis): Small groups are most effective at building knowledge if members share interests but bring to bear diverse backgrounds and perspectives.
  - H2 (collaborative curriculum hypothesis): Educational activities can be designed to encourage and structure effective collaborative learning by presenting open-ended problems requiring shared deep understanding.
  - H3 (collaborative technology hypothesis): Online computer support environments can be designed to facilitate effective collaborative learning that overcomes limitations of face-to-face communication.
  - H4 (collaborative cognition hypothesis): Members of collaborative small groups can internalize group knowledge as their own individual knowledge and they can externalize it in persistent artifacts.
  - H5 (collaborative methodology hypothesis): Quantitative and qualitative analysis and interpretation of interaction logs can make visible to researchers the online learning of small groups and individuals.

The project team of this multidisciplinary researcher/educator collaboration consists of: five co-PIs; curriculum teams consisting of student teachers, classroom teachers and Math Forum staff; technology teams including Math Forum technical staff and three graduate research assistants; and teams of national and international CSCL researchers assisting with methodology development and the study of group cognition. PI Stahl has developed numerous collaborative learning environments and has published on related issues in CSCW and CSCL theory and methodology. Co-PI Weimar has been Director of the Math Forum since its founding in 1994. Co-PI Bach is professor of educational technology. Co-PI Shumar is an educational ethnographer and long-time evaluator of the Math Forum. Co-PI Robertson is a cognitive scientist who teaches quantitative research methods and natural language processing. The Math Forum has a substantial, experienced staff. Some two dozen national and international researchers with relevant specialties are collaborating on this project.
Intellectual merit. This project addresses straight-on a central hypothesis of collaboration theory and CSCL that has never been adequately clarified and documented. It studies this issue in the context of extending the services of one of the most successful online educational institutions, simultaneously studying closely related issues of group composition, curriculum design, technology support and assessment methodology. The project brings together five co-PIs with the required mix of expertise, along with teams of engaged educators and researchers.

Broader impact. The project develops rigorous methods for studying a controversial issue in educational theory: the online construction of group knowledge that can exceed individual knowledge. It reconciles basic research and educational practice by providing a research base for CSCL and for virtual learning communities like the Math Forum. It develops public mechanisms to bring diverse individuals from around the Web together to learn.

1. A Central Research Issue of Collaborative Learning

1.1. A Conflict of Paradigms

Research on learning and education is troubled to its core by a conflict of paradigms. Sfard reviewed some of the history and consequences of this conflict in terms of the incompatibility of the acquisition metaphor (AM) of learning and the participation metaphor (PM) (Sfard, 1998). AM conceives of education as a transfer of knowledge commodities and their subsequent possession by individual minds. Accordingly, empirical research in this paradigm looks for evidence of learning in changes of mental contents of individual learners. PM, in contrast, locates learning in intersubjective, social or group processes, and views the learning of individuals in terms of their changing participation in the group interactions. AM and PM are as different as day and night, but Sfard argues that we must learn to live in both complementary metaphors.

The conflict is particularly pointed in the field of CSCL (computer-supported collaborative learning). The term —collaborative learning“ can itself be seen as self-contradictory given the tendency to construe learning as taking place in individual minds. Having emerged from a series of paradigm shifts in thinking about instructional technology (Koschmann, 1996), the field of CSCL is still enmeshed in the paradigm conflict between opposed cognitive and sociocultural focuses on the individual and the group
Proposals for Research

(Kaptelinin & Cole, 2002). In a keynote at the CSCL ‘02 conference (Stahl, 2002f), Koschmann argued that even exemplary instances of CSCL research tend to adopt a theoretical framework that is anathema to collaboration (Koschmann, 2002a). Koschmann recommended that talk about — knowledge“ as a thing that can be acquired should be replaced with discussion of —meaning-making in the context of joint activity“ in order to avoid misleading images of learning as mental acquisition and possession.

Although Koschmann’s alternative phrase can describe the intersubjective construction of shared meanings achieved through group interaction, the influence of AM can re-construe meaning-making as something that must perforce take place in individual human minds, because it is hard for most people to see how a group can possess mental contents. In a paper at CSCL ’03 responding to Koschmann’s earlier keynote, the PI argued that both Koschmann’s language and that of the researchers he critiqued is ambiguous and is subject to interpretation under either AM or PM (Stahl, 2003c). A simple substitution of wording is inadequate; it is necessary to make explicit when one is referring to individual subjective understanding and when one is referring to group intersubjective understanding œ and to make clear to those under the sway of AM how intersubjectivity is concretely possible.

The problem with recommending that researchers view learning under both AM and PM or that they be consistent in their theoretical framing is that our common sense metaphors and widespread folk theories are so subtly entrenched in our thinking and speaking. The languages of Western science reflect deep-seated assumptions that go back to the ideas of Plato’s Meno and the ego cogito of Descartes’ Meditations. It is hard for most people to imagine how a group can have knowledge, because we assume that knowledge is a substance that only minds can acquire or possess, and that only physically distinct individuals can have minds (somewhere in their physical heads).

1.2 Evidence to Overcome the Conflict

We (the PI, co-PIs and collaborators in this project) propose to address this central research issue head-on by studying online collaborative learning in the specific context of Math Forum problems, with the aim of presenting empirical examples of concrete situations in which groups can be seen to have knowledge that is distinct from the knowledge of the group members. By analyzing these situations in detail, we will uncover mechanisms by which understanding of mathematics passes back and forth between the group as the unit of analysis and individual group members as units of analysis.
One example might be a group of 5 high school students collaborating online over a two week period. They solve an involved algebra problem and submit a discussion of their solution to the Math Forum. By looking carefully at the computer logs of their interactions in which they collaboratively discussed, solved and reflected upon the problem, we can see that the group solution exceeds the knowledge of any individual group members before, during or after the collaboration. For instance, there may be some arguments that arose in group interaction that none of the students fully understood but that contributed to the solution. Or a mathematical derivation might be too complicated for any of the students to keep—in mind“ without reviewing preserved chat archives or using an external representation the group developed in an online whiteboard. By following the contributions of one member at a time, it may also be possible to find evidence of what that student understood before, during and after the collaboration, and thereby to follow individual trajectories of participation in which group and individual understandings influenced each other.

While we do not anticipate that group knowledge often exceeds that of all group members under generally prevailing conditions, we hypothesize that it can do so at least occasionally under particularly favorable conditions. We believe that we can set up naturalistic conditions as part of a Math Forum service and can collect sufficient relevant data to demonstrate this phenomenon in multiple cases. The analysis and presentation of these cases should help to overcome the AM/PM paradigm conflict by providing concrete illustrations of how knowledge can be built through group participation as distinct from œ but intertwined with œ individual acquisition of part of that knowledge. It should also help to clarify the theoretical framing of acts of meaning-making in the context of joint activity.

We believe that the theoretical confusion surrounding the possibility of group knowledge presents an enormous practical barrier to collaborative learning. Because students and teachers believe that learning is necessarily an individual matter, they find the effort at collaborative learning to be an unproductive nuisance. For researchers, too, the misunderstanding of collaborative learning distorts their conclusions, leading them to look for effects of pedagogical and technological innovation in the wrong places. If these people understood that groups can construct knowledge in ways that significantly exceed the sum of the individual contributions and that the power of group learning can feed back into individual learning, then we might start to see the real potential of collaborative learning realized on a broader scale. This project aims to produce rigorous and persuasive
empirical examples of collaborative learning to help bring about the necessary public shift in thinking.

1.3. The Range of Views on this Issue

CSCL grows out of research on cooperative learning that demonstrated the advantages for individual learning of working in groups (e.g., Johnson & Johnson, 1989). There is still considerable ambiguity or conflict about how the learning that takes place in contexts of joint activity should be conceptualized. While it has recently been argued that the key issues arise from ontological and epistemological commitments deriving from philosophy from Descartes to Hegel (Koschmann, 2002b; Packer & Goicoechea, 2000), we believe that it is more a matter of focus on the individual (cognitivist) versus group (sociocultural) as the unit of analysis (Stahl, 2003b, 2003c). Positions on the issue of the unit of learning take on values along a continuous spectrum from individual to group:

- Learning is always accomplished by individuals, but this individual learning can be assisted in settings of collaboration, where individuals can learn from each other.

- Learning is always accomplished by individuals, but individuals can learn in different ways in settings of collaboration, including learning how to collaborate.

- Groups can also learn, and they do so in different ways from individuals, but the knowledge generated must always be located in individual minds.

- Groups can construct knowledge that no one individual could have constructed alone by a synergistic effect that merges ideas from different individual perspectives.

- Groups construct knowledge that may not be in any individual minds, but may be interactively achieved in group discourse and may persist in physical or symbolic artifacts such as group jargon or texts or drawings.

- Group knowledge can be spread across people and artifacts; it is not reducible to the knowledge of any individual or the sum of individuals’ knowledge.

- All human learning is fundamentally social or collaborative; language is never private; meaning is intersubjective; knowledge is situated in culture and history.
Individual learning takes place by internalizing or externalizing knowledge that was already constructed inter-personally; even modes of individual thought have been internalized from communicative interactions with other people.

Learning is always a mix of individual & group processes; the analysis of learning should be done with both the individual and group as units of analysis and with consideration of the interplay between them.

In this project, we take a rather strong position on collaborative learning as our working hypothesis:

H0 (collaborative learning hypothesis): A small online group of learners can (on occasion and under favorable conditions) build knowledge and understanding that exceeds that of its individual members.

The different positions listed above are supported by a corresponding range of theories of human learning. Educational research on small group process in the 1950’s and ’60’s maintained a focus on the individual as learner (Johnson & Johnson, 1989; Stahl, 2000b). Classical cognitive science in the next period continued to view human cognition as primarily an individual matter œ internal symbol manipulation or computation across mental representations, with group effects treated as secondary boundary constraints (Simon, 1981; Vera & Simon, 1993). In reaction to these views, a number of sociocultural theories have become prominent in the learning sciences in recent decades. To a large extent, these theories have origins in much older works that conceptualized the situated-ness of people in practical activity within a shared world (Bakhtin, 1986; Heidegger, 1927/1996; Husserl, 1936/1989; Marx, 1867/1976; Schutz, 1967; Vygotsky, 1930/1978). Here are some representative theories that focus on the group as a possible unit of knowledge construction:

Collaborative Knowledge Building. A group can build knowledge that cannot be attributed to an individual or to a combination of individual contributions (Bereiter, 2002; Fuks, Gerosa, & Pereira de Lucena, 2001; Hakkarainen & Lipponen, 2002; Scardamalia & Bereiter, 1996; Wasson & Morch, 2000).

Social Psychology. One can and should study knowledge construction at both the individual and group unit of analysis, as well as studying the interactions between them (Daradounis, Xhafa, & Marques, 2003; Fischer & Granoo, 1995; Palen, 1999; Resnick, Levine, & Teasley, 1991).
Distributed Cognition. Knowledge can be spread across a group of people and the tools that they use to solve a problem (Hutchins, 1996; Hutchins & Palen, 1998; Solomon, 1993; Wasson & Morch, 2000).

Situated Cognition. Knowledge often consists of resources for practical activity in the world more than of rational propositions or mental representations (Hewitt, Scardamalia, & Webb, 1998; Polanyi, 1966; Schön, 1983; Suchman, 1987; Winograd & Flores, 1986).


Zone of Proximal Development. Children grow into the intellectual life of those around them; they develop in collaboration with adults or more capable peers (Brown & Campione, 1994; Goldman-Segall, 1998; Hmelo-Silver, 2004; Lemke, 1990; Vygotsky, 1930/1978).

Activity Theory. Human understanding is mediated not only by physical and symbolic artifacts, but also by the social division of labor and cultural practices (Engeström, 1999; Gay & Bennington, 1999; Kaptelinin, 1996; Nardi, 1996a; Nardi, 1996b).

Ethnomethodology. Human understanding, inter-personal relationships and social structures are achieved and reproduced interactionally (Dourish, 2001; Garfinkel, 1967; Hall, 1999; Heritage, 1984; Koschmann & LeBaron, 2003; Stahl, 2002d; Streeck, 1996; Streeck & Mehus, 2003).

One does not have to commit to one of these theories in particular in order to gain a sense from them of the possible nature of group knowledge. We have selected a working hypothesis that is in line with these theories in general without opting for one specifically. Based on our previous empirical work, we believe that we can study the issues raised by these theories without circularity by structuring collaborative activities, varying their parameters and critically evaluating the results. By reflecting on the theoretical issues within our work, we believe we can avoid the pitfalls of theory-laden research without claiming unattainable value neutrality.

1.4. Empirical Study of Group Knowledge

The PI previously conducted a pilot study involving a group of five middle school students collaborating on a problem involving data from a computer simulation. Like many studies of collaborative learning (e.g., Hmelo-Silver,
(but unlike the proposed study), this one involved face-to-face interaction with an adult mentor present. Close analysis of student utterances during an intense interaction suggested that the group developed an understanding that certainly could not be attributed to the utterances of any one student (Stahl, 2002d). In fact, the utterances themselves were meaningless if taken in isolation from the discourse and its activity context.

There were a number of limitations to the pilot study: (1) Although the mentor was quiet for the specific interaction analyzed, it might be possible to attribute something of the group knowledge to the mentor’s guiding presence. (2) The digital videotape was limited in capturing gaze and even some wording. (3) The data included only two sessions, too little to draw conclusions about how much individual students understood of the group knowledge before, during or after the interaction. To overcome such limitations, the proposed study will (1) not involve mentors active in the collaborative groups although the group will work on problems that have been carefully crafted to guide student inquiry and advice can be requested by email from Math Forum staff. (2) The online communication will be logged, so that researchers have a record of the complete problem-solving interaction. (3) Groups will be studied over a period of a couple weeks and longer for several groups that work on a sequence of problems.

Despite its limitations, the pilot study clearly suggests the feasibility of studying group knowledge. It shows that group knowledge is constructed in discourse and that discourse analysis can —make visible“ that knowledge to researchers. Student discourse is increasingly recognized as of central importance to science and math learning (Atkins, 1999; Bauersfeld, 1995; Lemke, 1990; Schifter, 1996). Discourse analysis is a rigorous human science, going under various names: conversation analysis, interaction analysis, micro-ethnography, ethnomethodology (Coulthard, 1977; Duranti, 1998; Garfinkel, 1967; Heritage, 1984; Jordan & Henderson, 1995; Mehan, 1979; Sacks, 1992; Sinclair & Coulthard, 1975; Streeck & Mehus, 2003).

The focus on discourse suggests a solution to the confusion between individual and group knowledge, and to the conceptual conflict about how there can be such a thing as group knowledge distinct from what is in the minds of individual group members (Stahl, 2003b). One way of putting it is that meaning is constructed in the group discourse. The status of this meaning as shared by the group members is itself something that must be continually achieved in the group interaction; frequently the shared status — breaks down“ and a —repair“ is necessary. In the pilot study, the interaction of interest centered on precisely such a repair of a breakdown in shared
understanding among the discussants (Stahl, 2002d). While meaning inheres in the discourse, the individual group members must construct their own interpretation of that meaning in an on-going way. Clearly, there are intimate relationships between the meanings and their interpretations, including the interpretation by one member of interpretations of other members. But it is also true that language can convey meanings that transcend the understandings of the speakers and hearers. It may be precisely through divergences among different interpretations or among various connotations of meaning that collaboration gains much of its creative power (Stahl, 2003c). These are questions that we will investigate as part of our micro-analytic studies of collaboration data, guided by our central working hypothesis. We believe that such an approach can maintain a focus on the ultimate potential in CSCL, rather than losing sight of the central phenomena of collaboration as a result of methods that focus exclusively on statistical trends (Stahl, 2002a).

1.5. Related Issues for Investigation

Collaborative success is hard to achieve and probably impossible to predict. CSCL represents a concerted attempt to overcome some of the barriers to collaborative success, like the difficulty of everyone in a group effectively communicating their ideas to all the other members, the complexity of keeping track of all the inter-connected ideas that have been offered or the barriers to working with people who are geographically distant. As appealing as the introduction of technological aids for communication, computation and memory seem, they inevitably introduce new problems, changing the social interactions, tasks and physical environment. Accordingly, CSCL study and design must take into careful consideration the social composition of groups, the collaborative activities and the technological supports.

In order to observe effective collaboration in an authentic educational setting, we will adapt a successful math education service to create conditions that will likely be favorable to the kind of interactions that we want to study. We must bring together groups of people who will work together well, both by getting along with and understanding each other and by contributing a healthy mix of different skills. We must also carefully design mathematics curriculum packages that lend themselves to the development and display of deep math understanding through collaborative interactions œ open-ended problems that will not be solved by one individual but that the group can chew on for a week or two of online interaction. Further, the technology that we provide to our groups must be
easy to use from the start, while meeting the communicative and representational needs of the activities. As part of our project, we will study how to accomplish these group formation, curriculum design and technology implementation requirements. This is expressed in three working hypotheses of the project: H1, H2 and H3 listed in section 3.2.

Two further working hypotheses define areas of knowledge building that the project itself will engage in on the basis of our findings. H4 draws conclusions about the interplay between group and individual knowledge, mediated by physical and symbolic artifacts that embody knowledge in persistent forms. H5 reports on the analytic methodology that emerges from the project.

2. The Math Forum Setting of Educational Practice

The Math Forum (www.mathforum.org) is a well-established digital library for mathematics education hosted at Drexel University. Started and periodically expanded with NSF funding (see section 3.4 below), the Math Forum now serves a virtual community of about one million people, including K-12, home-schooled and college students, classroom teachers, student teachers, mathematicians and life-long learners. It provides an integrated set of resources and services. The digital library consists of well over a million web pages of organized math resources that are publicly available. Math Forum staff and volunteers mentor thousands of students, and compile online summaries of advice given. There are special programs for students, teachers and student teachers. The second most popular service is the Problem of the Week. This service posts math problems at various grade levels (K-14) in core math subjects (e.g., algebra, geometry, pre-calculus) for students to work on. Students submit their solutions and a description of how they solved the problem to Math Forum; they receive feedback; the best submissions are posted. Mentoring is also available from the popular Ask Dr. Math service.

The Math Forum is an active virtual community, marked by strong currents of changing participation (Renninger, Weimar, & Klotz, 1989; Renninger & Shumar, 1998; Renninger & Shumar, 2002a, 2002b; Shumar & Renninger, 2002). A student with little interest in math may be introduced to the Math Forum through Problems of the Week integrated into her teacher’s lesson plans, and then be drawn in to explore further on her own. A person interested in math might start to participate in threaded discussions on the site, then do some mentoring and eventually contribute materials to the resource library. A programmer can download math applets, extend them
and contribute back. Many teachers and student teachers join teams to develop new Problems of the Week and attend Math Forum workshops to design new services.

Math Forum leadership is interested in significantly growing community activity and participation, so that people help each other, benefit from each other's knowledge and learn together. One way to do this is to mediate between the individual as solitary problem solver and the community as a whole by encouraging and supporting the formation of small groups of collaborators. Collaborative learning in small groups seems like the best way to reap the potential of the diverse virtual community of people of all ages and backgrounds interested in mathematics.

The proposed project is designed to systematically promote the formation of collaborative groups of math learners, based on rigorous research about group knowledge building. The project will first of all explore the ultimate potential of collaboration by studying the conditions under which group knowledge can be constructed that exceeds the knowledge of the group members. Secondarily, it will study the conditions that are favorable to successful collaboration, including group composition, curriculum design and technology support.

Individual student users of Math Forum will be invited to join small groups to work on special math problems, based on their previous activities at the Math Forum and their demonstrated skills. Teams of project staff and volunteers will develop Problems of the Week that are specially designed for collaborative learning by these groups. Guided by leading CSCL researchers, project staff will extend Math Forum technology to support the communication and representation needs of the student groups solving these problems collaboratively. Project staff, working with national and international CSCL researchers, will develop and use rigorous, reproducible, practical methods for analyzing the individual and group knowledge building that takes place in the collaborative groups.

3. Project Goal, Objectives, Approach, Prior Work

3.1. Goal

The central goal of the project is to investigate the following working hypothesis:
H0 (collaborative learning hypothesis): A small online group of learners can (on occasion and under favorable conditions) build knowledge and understanding that exceeds that of its individual members.

Although this is an empirical and falsifiable hypothesis, it is unlikely that it will be disconfirmed in this project. A failure to identify instances that confirm this hypothesis within the project data would probably just suggest that the project had not established sufficiently favorable conditions. Rather than trying to either simply confirm or disconfirm this hypothesis, the project aims to clarify the sense in which group knowledge can exceed member knowledge and show concrete cases that illustrate this elusive phenomenon.

3.2. Objectives

In addition to clarifying H0, the project’s objective is to determine conditions that are favorable to the construction of group knowledge. These conditions include the composition of the group (H1), the activities the group engages in (H2), and the technology (software artifacts, communication media) that mediates the collaborative interaction (H3). In particular, to understand how group knowledge can exceed member knowledge, it is important to analyze the relationship of group and individual knowledge and how they can transform into each other (H4). As a critical, reflective research project, the project also aims to specify a research methodology that is consistent with the project findings about the nature of group knowledge (H5). The project therefore studies the following working hypotheses:

H1 (collaborative group hypothesis): Small groups are most effective at building knowledge if members share interests but bring to bear diverse backgrounds and perspectives.

H2 (collaborative curriculum hypothesis): Educational activities can be designed to encourage and structure effective collaborative learning by presenting open-ended problems requiring shared deep understanding.

H3 (collaborative technology hypothesis): Online computer support environments can be designed to facilitate effective collaborative learning that overcomes limitations of face-to-face communication.

H4 (collaborative cognition hypothesis): Members of collaborative small groups can internalize group knowledge as their own individual knowledge and they can externalize it in persistent artifacts.
H5 (collaborative methodology hypothesis): Quantitative and qualitative analysis and interpretation of interaction logs can make visible to researchers the online learning of small groups and individuals.

3.3. Team Approach

The project studying small group collaboration will itself be conducted in a collaborative way, organized into five sets of interacting teams:

T1: Learning Teams: These are the small student groups that will be the primary subjects of the study. They will be groups of 3 to 5 students working together to solve special Math Forum Problems of the Week. Most Learning Teams will be formed online and will work online, generally never meeting each other face-to-face. During the first project year, however, some face-to-face groups of students in local schools will be studied to form a baseline for comparison purposes. Co-PI Weimar will coordinate the recruitment of Learning Team members.

T2: Curriculum Teams: Teams of student teachers, classroom teachers, mathematicians and Math Forum staff will work together to develop special Problems of the Week and associated curriculum for use by the Learning Teams. These Curriculum Teams will also be involved in analyzing the responses to the Problems they develop. Co-PI Bach will coordinate the work of the Curriculum Teams.

T3: Technology Teams: Math Forum technical staff and two project GRAs will work in teams under direct supervision by PI Stahl to design, implement and test collaboration software components. A user-centered design process will be followed, incorporating extensive user testing and iteration throughout the design cycle. Annual week-long workshops for the Curriculum, Cognition and Methodology Teams will provide focused opportunities for teachers and researchers with important experience and expertise to participate in the design process.

T4: Cognition Teams: The national and international researchers will be divided into Cognition Teams and Methodology Teams. The Cognition Teams will be supervised by co-PI Shumar and supported by one project GRA. They will analyze the computer logs of the online groups and any video data or other ethnographic data of the face-to-face Learning Teams.

T5: Methodology Teams: The Methodology Teams, supervised by co-PI Robertson, will assist in the quantitative analysis of data, and will refine and document the research methodology of the project.
3.4. Prior Work on the Math Forum

NSF REC-9155710, Geometry Forum, $1,439,238, June 1992 to November 1996
NSF REC-9618223, Math Forum, $2,922,166, March 1997 to February 29, 2000
NSF REC-9805289, BRAP, $293,924 (subcontract), August 1998 to February 2003
NSF DLI-2980185, JOMA, $651,948, July 2000 to September 2003
NSF NSDL 0085861, MathDL, $856,257, September 2000 to August 2003
NSF DUE 0127516, Online Mentoring Project, $137,300, February 2002 to January 2004
NSF NSDL 0226284, Math Tools, $853,113, September 2002 to August 2004

The Math Forum is arguably the most widely used math education site on the Internet (search for —math— on Google). It began in January of 1996 as a proof-of-concept grant from the NSF to extend the work of the Geometry Forum into other areas of mathematics and to investigate the viability of a virtual center for mathematics education on the Internet. The Math Forum has developed a vast Web site (mathforum.org) of over a million learning resources and it received more than 650,000 distinct visitors a month (making 2 million visits) in 2001, with mentored user services such as Ask Dr. Math, Problem of the Week, and Teacher2Teacher.

The Math Forum home page allows browsing and searching the Internet Mathematics Library of over 8,600 annotated entries of hand-selected resources. The cataloguing features are based on American Mathematical Society categories, and are enhanced by recommendations of the American Mathematics Metadata Task Force.

The Math Forum’s JOMA project (1) searched the Web and other resources to locate and collect applets and similar programs developed by the mathematics research and teaching communities, (2) reviewed and tested these systematically, and (3) made them easily accessible to undergraduate faculty and students. JOMA, the Journal of Online Mathematics and its Applications, is published by the Mathematical Association of America. The ESCOT project was a testbed for the integration of innovative technology in middle school mathematics. The Math Forum, working with SRI and other partners, developed team-based approaches that produced math tools for integration into the Problems of the Week. The JOME and ESCOT projects formed the basis for MathDL, an undergraduate-level digital library, a joint project of the MAA and the Math Forum, which is developing the technical infrastructure.
The MathDL and previous projects have given the Math Forum considerable experience constructing libraries and supporting technologies, such as metadata for the NSF digital library initiative. In addition, numerous Math Forum staff members have contributed to NSDL activities, meetings and working groups. The Math Forum was a founding member of the SMETE Open Federation, the largest identifiable user base for the National STEM Education DL.

With TERC and Michigan State University, the Math Forum investigated the possibilities for multimedia articles to open more effective communication between researchers and teachers in a project Bridging Research and Practice. The Math Forum developed a collaborative process through which teachers designed and conducted research into the use of discourse in the math classroom. A video-paper was produced jointly with researchers that served as the focal point for an online conversation with the mathematics education community at large.

Currently, the Online Mentoring project is developing a guide to enable professors to integrate online mentoring experiences into their mathematics and mathematics education courses. Pre-service teachers in these courses mentor students submitting their solutions to the Math Forum's Problems of the Week. The results of this project will be used to train mentors for — Technology PoWs,“ part of a new NSDL funded digital library of mathematics software.

The Math Forum provides many ways for people to interact with one another, with different points of access for people of varied strengths, needs and interests. Community building is an important part of Math Forum activities and has formed the basis of much of the content development on the site. The Math Forum represents a vision about the possibilities for an Internet community that extends the collegiality found in schools, classrooms or the workplace (Renninger & Shumar, 2002b). Evaluation of the Math Forum is used in program design, development, and facilitation, and provides an assessment of impact (Renninger et al., 1989; Renninger & Shumar, 1998).

3.5. Prior Work on Collaborative Learning

NSF CSS IRR-9711951, $725,000, September 1997 œ August 2000, — Conceptual Frameworks and Computational Support for Organizational Memories and Organizational Learning (OMOL),“ PIs: G. Fischer, G. Stahl, J. Ostwald

European Commission IST-2000-26249, May 2001 – May 2003, — Innovative Technology for Collaborative Learning,“ Fraunhofer-FIT (including G. Stahl) and researchers in Finland, Spain, Netherlands, Italy and Greece

The CSS grant for the OMOL project was instrumental in the PI’s turn from earlier work on organizational memory to support for collaborative learning. The project started from a model of computer support for organizations as domain-oriented design environments in which both domain knowledge and local knowledge are stored in the form of artifact designs and associated design rationale (Fischer et al., 1993). This CSCW model evolved into one of Collaborative Information Environments, that emphasized the interactive, asynchronous, persistent discussion of concepts and issues within an organization (Stahl, 2000a). Gradually, interest in organizational learning aspects led to involvement in CSCL and a model of collaborative knowledge-building environments (Stahl, 2001). A number of software prototypes were developed to explore the use of the Web as a communication and collaboration medium. Of these, the most important for the proposed work was WebGuide (Stahl, 2001), a prototype threaded discussion system that provided multiple perspectives on the discussion, comparison of perspectives and control over rearrangement of notes. Deployment of WebGuide in classrooms raised serious issues of adoption and concerns of socio-technical and social informatics (Kling, 1999) issues: motivation, media competition, critical mass, social practices, seeding, management, re-seeding, convergence of ideas, peer-to-peer collaboration, deployment strategies.

The EAR grant funded the initial implementation of WebGuide as an integrated Java applet supporting personal and group perspectives. It was a joint effort between the PI, a middle school teacher, and a research group at the NOAA labs in Boulder. The teacher taught an environmental science class in which he wanted to spend the year having his students interview various adults and construct a set of contrasting perspectives (conservationist, regulatory, business, community) on a particular local environmental issue that the students had previously been involved in. WebGuide was used by the students to collect notes on their interviews and to formulate personal and team perspectives on the issue. Results of this software trial were analyzed and presented at conferences (Stahl, 1999a, 1999b; Stahl & Herrmann, 1999; Stahl, 2001).
The European Commission grant supported software design and development of BSCL by researchers in Finland, Germany and Spain. The software was implemented as extensions of BSCW, a mature CSCW product used by 200,000 unique users since 1996 (Appelt, 1999). The PI went to work with the BSCW team at Fraunhofer-FIT near Bonn, Germany, for the first year of the project. He prototyped the BSCL innovations and published descriptions of them (Stahl, 2002e, 2003a). During its second year, the project is assessing the use of the new software in schools in Finland, Netherlands, Italy and Greece.

3.6. Our Current Related Work and Related Proposals

Math Forum staff periodically try out mechanisms to support small group collaboration on a small scale. They have provided chat services or encouraged face-to-face groups in classrooms to submit team responses to Problems. These trials generally produce immediate interest from the community, indicating that systematic support for small groups could have dramatic results in stimulating participation in the Math Forum and the associated community.

The PI is exploring small group formation approaches and innovative software functionality to support small group collaboration in his online courses. His students study software support for small group formation, collaborative knowledge negotiation and group use of digital libraries. Each of his HCI courses engages in user studies, software design and user testing of specific applications in these areas.

The PI and co-PIs of this proposal recognize that many research and technical issues related to this project require careful research and technology innovation that go well beyond the scope of the current grant proposal. They have therefore submitted an NSF ITR proposal for innovative technology to form and support collaborative small groups and an NSDL proposal for related research on digital library services to collaborative small groups. Particular co-PIs are also involved in other pending NSF proposals. These related projects œ if funded œ would be complementary to the project proposed here, but mutually independent. Although co-PI hours might have to be adjusted, there would be different Research Assistants and different goals, objectives and timetables. The present proposal provides the basic research and the necessary evidential basis for the other work. It aims to study in depth the collaborative interactions and the knowledge jointly constructed in collaboratively solving Math Forum problems, using for the project resources and technologies that are almost at hand. Parallel efforts at technology innovation and digital
library services in the other projects would allow this study to be refined and significantly extended.

4. Project Plan

4.1. Project Team

The project team consists of five co-PIs (in various schools of Drexel University see 4.1), collaborating researchers (national and international leaders in CSCL, CSCW and HCI see 4.2) and members of the Math Forum community (student teachers, classroom teachers, mathematicians and Math Forum staff see 4.3).

College of Information Science & Technology

Drexel University has a deep interest in and commitment to online collaborative learning. It has large online programs in information science, library science and nursing that use current technologies like Blackboard, WebCT and Lotus Notes and that make extensive use of group learning. Drexel has a long history of computer technology leadership as a former Institute of Technology, including being the first university to require entering undergraduates to have a PC and more recently being judged the — most wired” university according to Yahoo.

Drexel University’s College of Information Science and Technology is rated the #1 graduate school of library science information systems by US News and World Report (www.usnews.com/usnews/edu/grad/rankings/lib/brief/infsp3_brief.php). This interdisciplinary college offers online and campus-based undergraduate and graduate programs in computer science (e.g., HCI, databases, software engineering) and library science (including digital libraries).

PI Gerry Stahl is an Associate Professor in Drexel University’s College of Information Science and Technology. He brings a multidisciplinary background to the project, with PhD dissertations in philosophy/social theory and computer science/AI (Stahl, 1975, 1993b). He has developed a series of collaboration support systems: Hermes (Stahl, 1993a), WebNet (Stahl, 2000a), WebGuide (Stahl, 1999a, 1999b; Stahl & Herrmann, 1999; Stahl, 2001), BSCL (Stahl, 2002e, 2003a), and other educational software: Teachers Curriculum Assistant (Stahl, Sumner, & Owen, 1995; Stahl, Sumner, & Repenning, 1995) and State-the-Essence (Kintsch et al., 2000; Stahl & dePaula, 2001).
Stahl specializes in CSCL research, having published on CSCL theory (Stahl, 1993a, 1998, 2000b, 2002b, 2003b, 2003c) and the use of discourse analysis as an assessment methodology (Stahl & Sanusi, 2001; Stahl, 2002a, 2002c, 2002d). He was Program Chair of CSCL 2002 and Editor of the CSCL 2002 Proceedings (Stahl, 2002f). He is Workshop Chair of CSCL 2003, and Communications Chair and founding Board member of the International Society for the Learning Sciences (ISLS) (isls.org). He teaches online and in-class undergraduate and graduate courses on HCI, CSCL and CSCW at Drexel, using small group collaborative learning methods.

Co-PI Scott Robertson also teaches HCI courses at Drexel’s College of Information Science and Technology. Trained in cognitive science, Robertson teaches courses on natural language processing and quantitative research methods as well.

The Math Forum

Co-PI Stephen Weimar has directed the Math Forum since 1994. The Math Forum is now hosted at Drexel University. The Math Forum began in 1992 as the Geometry Forum at Swarthmore College, expanded to the Math Forum in 1996. It was funded in its development by the National Science Foundation, but has become largely self-sustaining in its stable services. It has become one of the most successful applications of the Internet to education through the development of interactive services that bridge the higher education, K-12, and industry communities. These services form the basis for a digital library that generates high quality mathematical content, supports student learning, integrates the benefits of technology with education, and is used for teacher professional development and pre-service teacher education. The Math Forum now comprises over 1.2 million pages of content, has over 2 million visits a month, receives up to 9,000 queries a month at its —Ask Dr. Math“ expert service, and mentored over 27,000 students during the 2000-2001 school year through its —Problem of the Week“ services. Among its current projects are two NSF grants, one focused on the use of online student mentoring programs in pre-service teacher education courses, and the other on the development of MathTools, a digital library for software in mathematics education from arithmetic to calculus.

Education & Ethnography

Drexel University has a School of Education and a Department of Culture & Communication, both of which are represented in this project. Co-PI Craig Bach is a professor in the School of Education, where he explores the use of technology in education, having developed several hypermedia presentations of topics in mathematics and philosophy. Co-PI Wesley
Shumar is a cultural anthropology professor in the Department of Culture & Communication, who specializes in educational anthropology and has conducted ethnographic studies of the Math Forum for many years.

4.2. National and International Collaborators

An important feature of this project is the involvement of leading national and international researchers. They bring expertise from a variety of relevant specialties and perspectives. Their participation will provide a natural means for sharing practical knowledge from Europe and the US as well as for disseminating the results of this project across the nation and globe. To ensure a strong cadre of collaborators, the following researchers have already expressed strong interest in participating in the project; others can join in the future:

The American collaborators are mostly experienced NSF grantees with specialties in particular aspects of CSCL: Geri Gay (Cornell), Ricki Goldman (NJIT), Cindy Hmelo-Silver (Rutgers), Christopher Hoadley (Penn State), Timothy Koschmann (Southern Illinois U), Bonnie Nardi (Irvine), Leysia Palen (Colorado U.), Linda Puliam (California State U.), Mark Schlager (SRI), Dan Suthers (Hawai‘i).

The international collaborators are renowned contributors to the theory and practice of CSCL: Wolfgang Appelt (Fraunhofer-FIT, Germany), Thanasis Daradoumis (Barcelona, Spain), Hugo Fuks (Rio, Brazil), Jörg Haake (Distance U, Germany), Kai Hakkarainen (Helsinki, Finland), Thomas Herrmann (Dortmund, Germany), Ulrich Hoppe (Duisburg, Germany), Jim Hewitt (Toronto, Canada), Victor Kaptelinin (Umea, Sweden), Anders Morch (Oslo, Norway), Wolfgang Prinz (Aachen, Germany), Barbara Wasson (Bergen, Norway), Volker Wulf (Siegen, Germany).

These individuals are established leaders in the HCI, CSCW and CSCL research communities, having made important contributions in theory, system design and assessment methodology. They all recognize the importance of collaboration, both in theory and in practice. See the Biographical Sketches section for brief descriptions of each.

The proposed NSF project builds on the work of the European ITCOLE project and its BSCL software (Appelt, 1999; Stahl, 2002e). The PI was the primary designer and prototyper of the BSCL software when he worked at Fraunhofer-FIT in Germany. The project with the Math Forum will involve close collaboration with the BSCW/BSCL team at FIT and has their full support. FIT will continue to support the BSCL code, making it available for free to educational institutions throughout the world. They will also provide
training to project staff who will be modifying the BSCL code. FIT has granted a five year developers license to the PI to work on extending BSCL as part of this project. Both Wolfgang Appelt, the BSCW/BSCL team manager, and Wolfgang Prinz, the director of the CSCW department at FIT, personally support the proposed project and its collaboration with FIT (see Supplementary Documentation).

The idea of automated support of group formation for workgroups in online learning is a research topic at the Distance University of Germany (Fern-Uni, Hagen). Jörg Haake, who has begun research on this topic (Haake et al., 2003; Wessner et al., 2002) will be a close collaborator with this project. Several recent publications by the national and international collaborators related to this project are included in the References Cited section (in bold face).

4.3. Management Plan

The PI, Stahl, will have primary responsibility for all aspects of the project. Weimar and Stahl will share project fiscal management, with accounting maintained by the Math Forum and Drexel University. Weimar and Stahl will share project staff management, recognizing that many staff are long-time employees of the Math Forum, contributing part-time.

The project Management Team consists of the five co-PIs, and will meet twice a month.

The project Staff consists of the five PIs, four Math Forum curricular staff, three Math Forum technical staff and three project graduate research assistants:

- G. Stahl, Information Science œ Project Manager and Technology Teams Coordinator
- S. Weimar, Math Forum œ Learning Teams Coordinator
- C. Bach, Education œ Curriculum Teams Coordinator
- W. Shumar, Anthropology -- Cognition Teams Coordinator
- S. Robertson, Information Science œ Methodology Teams Coordinator
- Underwood, Math Forum -- Math Forum Ask Dr. Math
- ○ Fetter, Math Forum -- Math Forum Problem of the Week
- K. Lasher, Math Forum -- Math Forum Problem of the Week
- S. Alejandre, Math Forum -- Math Forum Problem of the Week
The project Staff will hold monthly meetings at the Math Forum offices. These meetings will plan detailed project milestones and activities; review progress made according to the milestones; prepare for up-coming activities; review and revise the project plan; and make other decisions about the project as needed. Minutes of these meetings will be posted on the project website with other resources for review by the various teams, acting as project advisors.

Project management will be conducted following a collaborative model, in keeping with the philosophy of the project. Project activities will involve the collaborative teams, with project staff providing staff support and taking responsibility to ensure tasks are accomplished. Each set of teams will be coordinated by a co-PI: Weimar (Learning Teams), Bach (Curriculum Teams), Stahl (Technology Teams), Shumar (Cognition Teams) and Roberson (Methodology Teams). The project is by its nature assessment-centric, evaluating the nature of group learning under different conditions. In addition, the co-PIs will be responsible for various aspects of evaluating the project itself. Shumar will coordinate the experimental design, ethnographic investigation, and formative and summative assessment of the project. While Shumar will focus on qualitative assessment, Robertson will focus on quantitative assessment of project data and of the project itself. Stahl is responsible for coordinating and evaluating the software design and development; Smith for integration of software into the Math Forum site; Weimar for involvement of students and teachers, as well as integration of project activities with other Math Forum activities; and Bach for pedagogical aspects of the project.

Math Forum staff will participate in planning, design and facilitation of the Learner and Curriculum Teams. Math Forum staff will help with logistics, using their existing systems and networks of contacts. They will also help with hosting workshops for the teams as needed. Shumar will coordinate all data collection, and will focus the teams as needed for formative evaluation tasks. Stahl is responsible for project reports, including annual reports to NSF, culling from team summaries. Stahl, Bach, Robertson and Shumar will
prepare papers for conferences. Stahl and Weimar will be responsible for dissemination within the NSF and international research communities.

4.4. Infrastructure Technology

This project plans to adapt existing technologies as much as possible and to combine compatible software components into an integrated environment to support collaborative use of the Math Forum by small groups working together on the Internet. Useful components for supporting collaborative communication are available in various configurations and on different programming platforms. The only major component that has to be designed from scratch is a group formation component, and this has largely been worked out in the PI's recent courses. There do not seem to be any group formation components currently available, although the idea is not unprecedented (Swanson, 1964; Twidale & Nichols, 1998). Some organizations have explored systems for locating expertise within their staffs (Ackerman & McDonald, 1996); but the techniques for that do not transfer to the problem of finding people with matching interests using a digital library. There have been some experiments with social awareness, to display other people who are viewing the same web page at the same time (Graether & Prinz, 2001), but this hint is not enough to support group formation. A — group formation“ project in Japan matched learning theories (Inaba et al., 2000; Supnithi et al., 1999), but not people. A prototype for group formation in Germany allowed students who knew each other to self-select groups (Haake, Schuemmer, & Haake, 2003; Wessner & Pfister, 2001; Wessner, Dawabi, & Haake, 2002), but this approach does not scale to groups who do not know each other personally. A spin-off of this German research is being expanded and developed for distance education; the project will collaborate with Jörg Haake and associates through the Technology Teams (see section on International Collaboration). It will also collaborate with H. Ulrich Hoppe and Bonnie Nardi, who have both prominently argued for supporting small group collaboration for tasks like digital library search (Hoppe & Zhao, 1994; Nardi & O'Day, 1996).

The PI began exploring support for group formation while teaching an online HCI (Human-Computer Interaction) course for graduate students at Drexel. His students studied the issue and came up with several low-fidelity prototypes that they subjected to user testing. The PI developed an automated grouping agent, which he used to form work groups in subsequent courses. In both the student prototypes and the grouping agent, groups were formed based on specific criteria about the participants: their schedules, their interests and their skill levels. These pilot studies for the
Proposed project suggest the kinds of balance that should be sought in forming distributed groups. For instance, if synchronous communication is to be possible within the group especially given different global time zones, members must have compatible schedules. On the other hand, collaborative teams often work best when there is a diversity of perspectives and skills, along with a commonality of interests. Thus, a matching algorithm must optimize certain similarities and other differences. Various theories of collaboration stress the power of heterogeneity, of the utility of seeing things differently: cognitive dissonance (Festinger, 1957), perspectives (Boland & Tenkasi, 1995; Goldnam-Segall, 1998; Stahl & Herrmann, 1999), interdependence (Johnson & Johnson, 1989), zone of proximal development (Vygotsky, 1930/1978), cognitive flexibility (Feltovich et al., 1996).

The PI has informally explored support for small groups in his classes using two different online collaboration environments: Blackboard and BSCL. Blackboard is a commercial system to support collaboration. It is used widely in university courses, particularly in the US. Blackboard can be extended (in Java) by third party developers using the Blackboard Building Blocks SDK (see buildingblocks.blackboard.com/bin/bbdn_info.pl).

BSCL (Basic System for Collaborative Learning) is a system with collaboration support for classrooms that is similar to Blackboard (Stahl, 2002d). It was designed and developed by the PI and others in 2001/2002 as part of a European Union research project. BSCL is an extension (developed in Python) to BSCW (Appelt & Klöckner, 1999), a shared repository CSCW system widely used in European research and learning organizations. It is available for free to academic organizations. The PI has a license to develop it during the period of this project (see Letter of Support in Supplementary Documentation).

The Math Forum already has an infrastructure of custom software (developed in an object-oriented Perl-based environment) to support the virtual community and digital library of math resources and activities. It is possible to extend this system in various directions, such as using ZOPE or other Open Source components, extending Blackboard or adapting features of BSCL. Java applets can also be developed, adapting from the PI’s WebGuide system. The project will select one of these approaches during its early phases.
4.5. Timeline

The three year project period is planned to be September 1, 2004 to August 31, 2007. Roughly, work during these years will be focused as follows, based on Drexel University's quarter calendar:

<table>
<thead>
<tr>
<th>Quarter</th>
<th>Learning Teams</th>
<th>Curriculum Teams</th>
<th>Technology Teams</th>
<th>Cognition &amp; Methodology Teams</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Fall 2004</td>
<td>Recruit teachers</td>
<td>Select appropriate existing PoWs</td>
<td>Hire 2 GRAs</td>
<td>Hire GRA; workshop after CSCL '04</td>
</tr>
<tr>
<td>2. Winter 2005</td>
<td>Start up F2F groups</td>
<td>Workshop for teachers</td>
<td>Review existing CSCL technology</td>
<td>Propose method</td>
</tr>
<tr>
<td>3. Spring 2005</td>
<td>Observe 3 F2F groups</td>
<td>Define curriculum approach</td>
<td>Study MF technology</td>
<td>Select log segments</td>
</tr>
<tr>
<td>4. Summer 2005</td>
<td>Observe 3 more F2F groups</td>
<td>Design 6 new PoWs</td>
<td>Study BSCL technology</td>
<td>Workshop after CSCL '05</td>
</tr>
<tr>
<td>5. Fall 2005</td>
<td>Form online groups</td>
<td>Design 6 new PoWs</td>
<td>Study Blackboard technology</td>
<td>Interpret 3 logs</td>
</tr>
<tr>
<td>6. Winter 2006</td>
<td>Start collaborative PoWs</td>
<td>Workshop for teachers</td>
<td>User requirements</td>
<td>Analyze cognition in logs</td>
</tr>
<tr>
<td>7. Spring 2006</td>
<td>Observe 5 online groups</td>
<td>Study student solutions</td>
<td>Software version 1</td>
<td>Prepare 1 paper</td>
</tr>
<tr>
<td>8. Summer 2006</td>
<td>Change group criteria</td>
<td>Re-design curriculum approach</td>
<td>User testing</td>
<td>Workshop after ICLS '06</td>
</tr>
<tr>
<td>9. Fall 2006</td>
<td>Observe 5 more online groups</td>
<td>Select 12 new PoWs</td>
<td>Release version 1</td>
<td>Revise method</td>
</tr>
<tr>
<td>10. Winter 2006</td>
<td>Observe 3 longitudinal groups</td>
<td>Workshop for teachers</td>
<td>Revise version 2</td>
<td>Interpret 5 logs</td>
</tr>
<tr>
<td>11. Spring 2007</td>
<td>Release group formation support</td>
<td>Write teaching manual</td>
<td>Release version 2</td>
<td>Analyze cognition in logs</td>
</tr>
<tr>
<td>12. Summer 2007</td>
<td>Review group formation results</td>
<td>Organize digital library of PoWs</td>
<td>Document software</td>
<td>3 papers &amp; handbook</td>
</tr>
</tbody>
</table>

4.6. Project Artifacts

The project will produce a number of concrete artifacts to make the knowledge gained by the project team persistent and available to others:

**Evidentiary Base of Interpreted Logs.** The Cognition Teams will produce a set of at least eight interpreted logs, illustrating the analysis of group knowledge building.

**Math Forum Small Group Services.** The project will produce an on-going service at the Math Forum (MF), supporting collaborative knowledge building in Learning Teams with appropriate Problems of the Week (PoWs).

**Group Formation Criteria.** The project will produce a published set of group formation criteria for the formation of effective Learning Teams.
Problems of the Week for Collaborative Learning. The Curriculum Teams will produce a digital library of Problems of the Week and associated curriculum appropriate for small group collaboration.

Collaboration Technology Design. The Technology Teams will produce designs and documentation for software components to support small group collaboration at the Math Forum.

Theory of Collaborative Knowledge Building. The Cognition Teams will produce at least three conference or journal papers drawing implications from the project for the theory of collaborative learning.

Methodology Handbook. The Methodology Teams will produce a methodology handbook documenting the approach they developed to the analysis of group and individual learning from computer logs of collaborative interactions.

4.7. Project Evaluation

The project as a whole will be considered successful if it achieves the goals and objectives stated in Section 3. That is, if it clarifies what it means to claim that the knowledge of a group can exceed that of its members and if it sheds light on the conditions that favor such collaborative learning. In more specific and measurable terms, the project will be considered successful if it keeps up with the milestones for each set of teams listed in 4.5 above and produces the set of artifacts described in 4.6. In addition, project staff will produce required NSF project reports and participate in associated NSF events.

By its design, the project incorporates in project activities the dissemination of training, research insights and practical tools:

- involving many students, graduate students, student teachers, teachers and researchers;
- integrating research with the educational practices of the Math Forum;
- developing academic papers and practical handbooks

Evaluation is a way of life for most members of the project team. The PI and co-PIs specialize in HCI, ethnography and education œ fields that emphasize evaluation of various kinds; Math Forum staff have participated in many NSF grant programs in which evaluation played a central role; the group of national and international researchers includes renowned experts in qualitative and quantitative evaluation.
Evaluation is central to the project. The main project activity is the evaluation of learning at both the group and individual unit of analysis (H0). There will also be evaluation of the effect of different conditions on this learning, including group composition (H1), curricular activities (H2) and collaboration technology (H3). At a more detailed level, the forms of interplay between the group and individual learning will be evaluated, partially in terms of internalization and externalization processes (H4). This will lead to evaluation and refinement of the project's evaluation methodology itself (H5). Following is an overview of the plan for these evaluation foci, subject to iterative revision as a result of on-going project results:

H0: During the first year of the project, 6 face-to-face groups will be observed working on Math Forum Problems of the Week collaboratively. Observations will take place in local school classrooms affiliated with the Math Forum and Drexel's School of Education, using digital videotaping and ethnographic observation. The primary analysis tool will be discourse analysis, but other techniques will be used to help define the individual and group learning (Fischer & Granoo, 1995). In these cases it will be possible to use pre- and post-tests of individual and group understanding, as well as talk-aloud protocols in individual sessions. This will provide the researchers with a baseline analysis of the relationship of individual and group knowledge. During the second and third years, at least 5 online groups will be studied each year. Of these, students from 3 groups will be observed longitudinally across a series of problems. The longitudinal data will be analyzed to see persistence of knowledge at the individual and group units. The longitudinal data will also shed light on issues of leadership, motivation and interest. Data from the online groups will consist primarily of computer logs of their interactions. There will also be questionnaires about their backgrounds prior to group formation. In a limited number of cases there will be pre- and post-tests of mathematical understanding and personal interviews. Again for the online groups, the primary analysis tool will be discourse analysis of logs. A project GRA will collect, and log all interaction data, with a brief description of each minute of discussion. Interactions will be coded, using a variation of the coding schemes developed by our collaborators (Hakkarainen & Lipponen, 2002; Hmelo-Silver, 2004). The logs and coding will help researchers to locate important interactions for the evaluation of learning at the group and individual unit. Several cases will be selected for intensive discourse analysis to determine if there is plausible evidence that the group knowledge and understanding exceeded that of the individual group members and to determine how that
took place e.g., by the generative power of partially understood words or the merging of perspectives.

H1: Different algorithms for matching students in small groups will be experimentally manipulated in both face-to-face and online Learning Teams. The research literature indicates both that diversity of perspective increases the power of collaboration and that commonality of grounding is necessary. By varying the mix of participants and assessing their success as measured by their problem solutions and their displayed depth of group understanding, the project will evaluate this tradeoff and seek an optimal balance. This working hypothesis is based on rules of thumb from face-to-face collaboration as well as the PI’s experience with online problem-based courses; online interaction differs from face-to-face in manifold subtle ways, requiring careful empirical study. Evaluation of success will combine measures of success in group process, quality of knowledge building and depth of understanding as reflected in logs and problem solution (Daradounis et al., 2003; Hakkarainen & Lipponen, 2002).

H2: The Math Forum has extensive experience and success in developing problems at different grade levels and in different areas of mathematics for solution by individuals. In this project, such problems will be adapted for small group collaboration. These collaborative problems may have to be more complex, involve more deep comprehension of mathematical principles, require the use of tools or representations and take longer than a week to discuss and solve. The development of a library of such problems will be challenging and will require both trial-and-error and careful evaluation of how Learning Teams actually do with them. Problems will be rated by the quality of collaborative interaction, the depth of math discussion that they generate, and other standard Math Forum criteria.

H3: There are a variety of CSCL software components available for incorporation or adaptation to this project. The Technology Teams will evaluate the applicability of these to the Math Forum context in the first year. Then they will specify, design, implement and test iterative versions of support for the Learning Teams. The software will be carefully evaluated for its fit in this particular, concrete educational practice, using HCI evaluation methods and analysis of communication and collaboration breakdowns in the logs.

H4: Interactions that were found under the H0 evaluation to indicate group knowledge exceeding member knowledge will then be interpreted through further discourse analysis to determine the extent to which group knowledge was either internalized by certain individuals or externalized in the group report.
H5: The ability of the project to accomplish the evaluations associated with H0 and H4 in particular will reflect the extent to which H5 is confirmed. As with H0, it is not so much a matter of verifying or falsifying the hypothesis as it is of the Methodology Teams of researchers refining the use of discourse analysis of computer logs to make visible the group and individual knowing that was expressed in the online interactions.

5. Anticipated Results & Impact

5.1. Dissemination & Outcomes

Dissemination of project results, both in the US and in Europe, is built into the project design. Dissemination to the international research community, to practicing educators and to the public generally will take place primarily through the following mechanisms:

**Involvement of international researchers.** Approximately two dozen researchers will be intimately involved in this project, primarily through the Cognition Teams and Methodology Teams. Many of their graduate students will also be involved. Most of these researchers are active in important research centers.

**Workshops at international conferences.** The project will sponsor one workshop per year to bring together international and American researchers collaborating with the project. This may be coordinated with international conferences on education such as CSCL, ICLS, CSCW, ECSCW, AERA and EARLI. Most of the researchers involved in this project regularly attend these conferences and present at them. These conferences will be primary sites for the presentation of results from this project. Project staff will also submit papers and organize presentations about the project results at these conferences.

**Involvement of teachers and student teachers.** Perhaps two dozen teachers and student teachers will be intimately involved in this project, primarily through the Curriculum Teams. As the results of this project become part of Math Forum’s regular services, increasing numbers of teachers and student teachers will participate in spontaneously formed curricular workgroups. The project will bring Curriculum Team members together in intensive week-long workshops each year.

**The Math Forum virtual community.** This is a rapidly growing community that already numbers over a million distinct individuals. They will learn about the results of this project as collaborative problems become
a regular feature of the Math Forum and as community participants are automatically invited to join small groups for collaborative learning of mathematics.

The NSF community. Project findings will be presented at NSF gatherings where Math Forum or project participants are involved.

5.2. Sustainability & Contribution

The results of this project, particularly the collaborative Problem of the Week service, will be fully incorporated in the Math Forum. The Math Forum is a permanent program within Drexel University, so that services developed in this project will continue to exist and to be used indefinitely. Although the Math Forum receives grants to engage in research and service expansion, it strives to develop revenue sources to sustain existing services. The collaboration services of this project will contribute to building new lines of revenue, including contracted services with school districts for which Math Forum will provide custom collaboration services and support.

5.3. Integration of Research & Education

The project itself integrates research and education. It provides empirical support for the design and development of resources and services to support math education over a broad range of school grades, as well as for meeting educational needs of employees, mathematicians and lifelong learners. It does this from a position situated in the practical work of the Math Forum. The specific content of this project applies technologies at the forefront of CSCL and CSCW research directly to educational practice. By implementing collaborative learning services in an established educational context, the project studies actual instances of the practice it hopes to foster.

5.4. Integrating Diversity

A central project hypothesis is that groups integrating specific kinds of diversity learn better. Project software will be designed to optimize diversity during the group formation process. The online environment removes many of the barriers to collaboration among people with differences in age, knowledge, gender, culture, physical abilities, geographic location. The project will demonstrate the creative power of diversity takes in collaboration.
5.5. Intellectual Merit

This project addresses straight-on a central hypothesis of CSCL that has never been adequately clarified and documented. It seeks to develop an evidentiary basis for a controversial assumption common in learning theories today: that group knowledge can be something distinct from the sum or overlap of the individual knowledge of the group members. This assumption lies at the center of much of the current debate and confusion in educational theory, yet little in the way of clear and incontrovertible empirical examples have been presented before now. This project studies the building of group knowledge in the context of extending the services of one of the most successful online educational institutions, simultaneously studying issues of group composition, curriculum design, technology support and assessment methodology. To accomplish this, the project brings together five co-PIs with the required mix of expertise, along with teams of engaged educators and researchers.

5.6. Broader Impacts

The project develops rigorous methods for studying the online construction of group knowledge. By overcoming the paradigm conflict concerning collaborative learning, the project reconciles basic research and educational practice and provides a research base for CSCL and for virtual learning communities like the Math Forum user community. It develops technical, social and pedagogical mechanisms to bring diverse individuals together to learn, and offers a model, support tools and a research base for collaborative educational services. It does this with the help of leading national and international educational researchers, promoting in-depth collaboration at this level as well. These are all contributions to realizing the potential of Internet technology to not only make human knowledge accessible to everyone, but to promote collaborative knowledge building on a global scale.

References


Fuks, H., Gerosa, M. A., & Pereira de Lucena, C. J. (2001). Use of categorization and structuring of messages in order to organize the discussion and reduce information overload in asynchronous textual
communication tools. Paper presented at the International Workshop on Groupware (CRIWG '01), Darmstadt, Germany.


technologically supported environments. Heidelberg, Germany: Springer Verlag.


Foundations: Educational Online Communities for At-Risk Youth

**Submitters:** Mike Atwood (Drexel/IST), Gerry Stahl (Drexel/IST), Fran Cornelius (Drexel/Nursing), Steve Weimar (Drexel/Math Forum), Debra McGrath (Drexel/Nursing, Inst for Healthcare Informatics)

**For submission to:** Philadelphia-area and national foundations

**Date:** Draft of January 9, 2011

**Introduction**

Drexel University is a leader in the development, deployment and adoption of telecommunications technology to support education. In particular, the College of Information Science and Technology (IST) and the College of Nursing and Health Professions (CNHP) operate major online curricula for college and graduate study. The Math Forum has pioneered successful use of networked digital libraries of educational materials for K-12 students.

The John C. Ford Program based in Houston, Texas, has introduced a successful tele-community educational after-school program in partnership with community schools. The intent of this proposal is to create a similar partnership with the parents and children attending the community schools of lower North Philadelphia, IST, CNHP, Math Forum and the Ford Program to create an educational online community for at-risk youth. The purpose of this project is to improve academic performance, life skills and health status of persons living in lower North Philadelphia.

**The Need**

The lower North Philadelphia community served by this project – known as the 11th Street corridor – is a community in need. There is a remarkably high concentration of public housing in this community: six conventional public housing developments with 5,583 residents. The population is
predominately African-American women heads of household and their children. There is a disparately high percentage of unemployed individuals and of families in poverty compared to other neighborhoods in Philadelphia. Families living in this community have the lowest median family and household income in the city. (See Appendix A.)

National data suggest that the cycle of poverty, poor health status, and low educational achievement has become self-perpetuating in many communities. As in other impoverished communities, the residents of 11th Street suffer from significant health issues. This community experiences higher rates of morbidity and mortality compared to other areas, especially due to diseases such as hypertension, diabetes, asthma, and high risk behaviors such as cigarette smoking, substance abuse and risky sexual behaviors. (See Appendix B.)

Education – the acquisition and application of knowledge – is a critical component in breaking this cycle. The Ford Program offers an opportunity to engage in a hands-on interactive learning environment that empowers, educates and enriches the lives of students and their families. Drexel University proposes an expansion of the Ford Program, which will offer an innovative collaborative intervention with broad-based impact.

Long-Range Goals

We will provide a unique and innovative after-school program for youth and their families in low-income neighborhoods of Philadelphia. Our broad goal is to provide an educational community in which students and their families will be able to develop competencies that will translate into the following:

- Students will develop the skills to succeed academically and become connected to professionals from the health and information sciences. Students will be able to explore these professions and enter career paths at an early stage, leading to reduced school dropout rates and increased college attendance.

- Parents will develop personal relationship and technical skills essential for the workplace. These are the tools needed for upward mobility in the work world.

- Both students and parents will join a community of learners and experts to acquire business, presentation and computer literacy skills. The acquisition of these skills will lead to a higher level of self-confidence and ultimately to an increased ability to advocate for self and family.
Partners

CNHP: The CHNP has been working with the target community for seven years and has a proven track record for successful programs in the targeted community in close collaboration with community leaders and the local school district.

Math Forum: The Math Forum will adapt its award-winning online services to help participating students develop their mathematical problem-solving skills and reinforce some of the math concepts and techniques found in the project curriculum. These online services and resources will provide continuity for the students and academic support between project events.

IST: Drexel IST will research technology transfer procedures for adapting software used at the university to local after-school settings in low-income Philadelphia neighborhoods, and will provide on-going research, development and evaluation services, partnering where appropriate with the School of Education and the Math & Computer Science Department.

John C. Ford Program: The Ford Program is a successful educational initiative that involves a unique blend of elements: neighborhood-based Inner-City Telecommunication Centers; a non-academic, real-world business format and curriculum with a focus on science, math, business and technology; strong support from corporate partners; a multimedia network that uses interactive videoconferencing with an innovative protocol to attract, engage and train low-income youth and adults; training and education for the whole family in low-income communities; and state-of-the-art technology that allows the program to tap into available learning resources. This program is now gearing-up for a national scaling initiative for dissemination.

Pilot Implementation

Curriculum: We will begin in Winter 2003 with after-school programs at Harrison Community Center in lower North Philadelphia to capitalize on the national scaling effort of the John C. Ford Program’s Global Tele-Communities Education Initiative. This one-year phase will use curriculum already proven successful in low-income neighborhoods in Houston. We will adapt the Ford Program curriculum – which focuses on science and business skills – for our target population. Initial offerings will include: “Science, Math and Technology” curriculum developed in cooperation with the Ford Program, Space Center Houston and the Math Forum. Other
offerings will be pre-SAT “Language Arts” and “Math Crafts” curricula from the Ford Program.

In the second year, we will offer “Healthy Habits”, a health literacy and self-efficacy curriculum developed by the College of Nursing. The prototype for this curriculum is currently under development as part of two studies in the College of Nursing; the HTN Study and the Asthma Education Program. The pilot project will provide needed experience and outcome data to enable us to design a more complete program and to seek federal funding.

**Mentoring:** Through telecommunicated simulations and online interactions students will receive learning support from experts in the field and higher education students and faculty, thus forming relationships that encourage students to expand their horizons and set higher expectations for academic performance.

**Modeling:** The business and science curricula enable the program to model problem-solving, teamwork and other strategies for success in academic and work environments.

**Recruitment:** We plan to publicize this opportunity through schools, the local community center and guidance counselors. The program will be open to all interested individuals, on a first-come, first served basis. There will be clear behavior and performance expectations with clear consequences and replacement strategies.

**Resources:** The Harrison Community Center, located in lower North Philadelphia public housing, is a hub for community activity. The center is operated by a very active and deeply committed resident council which strongly supports this initiative. The council has, for the past seven years, sought out opportunities and partnerships that encourage and support families in the pursuit of education and training as a means to a better life. The Harrison Computer Resource Center (HCRC), a modest computer lab, was established to provide area residents access to educational opportunities and to build skills for jobs that provide a living wage. The HCRC provides a vital service to area residents in providing access to technology, which is now considered a ‘life-skill’ essential for success in today’s work and school environments. The commitment by the community and community leaders make the HCRC a logical partner in this project.

**Evaluation:** We will assess effectiveness of the Pilot Implementation phase from data collected using the following methods:

- Youth Risk Inventory administered as the program begins and at the end.
Proposals for Research

❖ Pre- and post-inventories measuring attitudes toward substance abuse and other risky behaviors, school, work, and goal setting.

❖ Administration of inventories at the start of the program and at the end that assess and monitor self-efficacy, health status, and health behaviors.

❖ Monitoring of computer-based activities.

❖ Analysis of computer interaction logs.

❖ Student school performance (report cards, attendance).

❖ Program participation and retention records.

❖ Parent/guardian involvement records.

Scaling Up Process

As university educators and researchers, our priorities include the involvement of our own university students to:

❖ Develop a workable curriculum that involves at-risk youth and their families in developing self-efficacy.

❖ Develop a workable curriculum that involves at-risk youth and their families in developing healthy attitudes and knowledge about risky behaviors such as substance abuse, cigarette smoking, poor diet, early engagement in sex, etc.

❖ Develop a workable technological infrastructure to be used in low-income neighborhoods for learning and sharing healthy attitudes.

❖ Develop pedagogical and curricular approaches that are successful at involving at-risk youth and their families in educational online communities.

As we succeed in these areas, we will increase the number of sites in low-income Philadelphia neighborhoods offering after-school online educational communities and we will diversify the curriculum offerings at these after-school programs.

Request

We are requesting $88,000 for calendar year 2003. This will cover faculty and teacher release time for staffing the after-school program and small
amounts for supplies and overhead. Because we need to start up the Pilot Implementation program in early 2003 to coordinate with the Ford Program, there is insufficient time to request federal funding. Furthermore, the pilot will provide us with extensive hands-on experience setting up and working with this type of after-school program to support a major research and implementation proposal. Foundation funding for an initial pilot project year will thereby leverage substantial longer-term funding that can have a sizable impact on the at-risk population in Philadelphia.

**Budget for one-year pilot project**

$28,000 Project Management (20% release time for McGrath & Cornelius)

$24,000 Project Instruction (2 certified teachers * 10 hrs/wk * 30 wks * $40/hr)

$6,000 Software

$6,000 Hardware

$5,000 Travel (to Houston, etc.)

$5,000 Supplies

$14,000 Indirect (20%)

$88,000 Total

**Supporting Documents**

Letter of support from after-school program

Letters of support from local schools

Letter of support from the John C. Ford Program

One-page biographies of Principal Investigators
Part III: Grants Awarded at the University of Colorado
New Media to Support Collaborative Knowledge Building: Beyond Consumption and Chat

Executive Summary

The Center for LifeLong Learning and Design (L3D), directed by Gerhard Fischer, is an interdisciplinary research and teaching center within the Department of Computer Science and the Institute for Cognitive Science. It has long focused on developing theoretical frameworks and prototype technologies for supporting high-functionality, collaborative, creative tasks with computational media. In recent years, it has explored models of Web-based communication that foster interaction and the collaborative construction of shared knowledge.

The proposed project is situated within an advanced interdisciplinary seminar that brings together faculty and graduate students to research the role of computational cognitive artifacts and innovative Web-based media in collaborative learning and education. Much of the seminar activity takes place through WEBGUIDE, a knowledge-building environment being developed by Gerry Stahl, the project Principal Investigator. WEBGUIDE is an experiment in dynamically structuring hypertext communication according to group and personal perspectives.

The requested seed grant will allow the further development and evaluation of WEBGUIDE during the duration of the seminar. WEBGUIDE will become a central theme of the seminar as an example of a computational artifact that seminar members learn to use. New functionality will be added to WEBGUIDE incrementally as suggested by its use. The adoption, usability, benefits and limitations of the technology will be evaluated and reported.

L3D has always pursued interaction with industry. WEBGUIDE and related models and technologies developed at L3D have been used in a number of classroom settings, and will soon be ready for exploration in commercial settings.
Technical Description

The L3D Center

The Center for LifeLong Learning and Design (L3D) emerged from Gerhard Fischer’s research group on Human-Computer Communication. That research group established a reputation for developing ideas and systems to support high-functionality, collaborative, creative tasks with computational media. A central theme in the late 1980’s and early 1990’s was computational design environments to support designers in various fields (e.g., design of kitchens, voice-dialog systems, LANs, habitats in outer space. Most current L3D faculty were involved in these efforts, and approaches from it influence on-going research.

An important focus in the early work was on supporting “lifelong learning,” “just-in-time learning,” or “learning-on-demand”. The design environments were always “domain-oriented,” that is, they were built around knowledge-bases of domain knowledge. A distinctive feature of L3D’s research was a concern with the evolution of these knowledge-bases through use. It was seen as critical that designers not only have access to domain knowledge as they work, but that they can grow the knowledge-base by contributing to it and reorganizing it. This concern broadened the research interests from HCI (human-computer interaction) and AI (artificial intelligence) to CSCW (computer supported cooperative work) and CSCL (computer supported collaborative learning) (Arias et al., 1999; 2000; Fischer, 1994; 1998).

With the advent of the Web, L3D research shifted to exploiting the promise of the Web as an infrastructure for sharing knowledge and collaborating on design. Here it was important to distinguish distinct models of Web use: (model 1) the Web as a read-only repository of information; (model 2) the Web as a place where information may be submitted to webmasters who mediate its dissemination; (model 3) the Web as a communication medium in which users interactively grow shared knowledge. Of course, each of these models is appropriate for certain classes of use and model 3 raises a variety of special issues which we are currently investigating.

Much of L3D’s work recently has tried to identify limitations of popular models of the Web and to explore ways of overcoming these limitations. To allow users to move from the role of passive consumers of information to active producers of shared knowledge, we developed a number of prototypes and then deployed them in courses. For instance, a series of interlocking dynamic websites were created in which users could interactively build glossaries of technical terms, bibliographies of literature sources and
threaded discussion of topics (see DYNAGLOSS, DYNASOURCE, DYNACLASS: http://www.cs.colorado.edu/~ostwald/dynasites.html ).

We observed through analysis of entries in threaded discussions that these media, though interactive, were generally limited to relatively superficial chat or exchange of personal opinions. WEBGUIDE was developed to explore support for activities of knowledge-building that go beyond both consumption and chat to dialog, merging of perspectives, clarification of meanings, theory building, shared knowledge, crystallization of ideas in cultural artifacts, etc. (Stahl, 1999b; 1999c; 2000b).

Other L3D projects address the limitations of purely digital communication by allowing interaction with physical objects that have computational implications. Another major concern is that today’s technology excludes both people without access to equipment and also many people with physical or mental disabilities. L3D has begun a major effort – in association with industry – to develop technologies that address the special needs of these populations.

The Project Environment: Teaching & Research

The proposed project is situated within an advanced interdisciplinary seminar on “Perspectives in Computer Supported Collaborative Learning” (http://www.cs.colorado.edu/~gerry/readings/ ). The seminar brings together faculty and graduate students from Computer Science, Communication, Education, Psychology and Philosophy to research the role of computational cognitive artifacts and innovative Web-based media in collaborative learning and education. A variety of theoretical approaches are reviewed through discussion of seminal texts and collaborative micro-ethnographic analysis of videos from a middle school classroom. Video clips, log and transcripts are available through WEBGUIDE, which also provides the medium for communication and group theory building. Several seminar participants interact solely through WEBGUIDE from other universities and even from other countries.

WEBGUIDE is an experiment in structuring hypertext communication according to group and personal perspectives. Seminar participants each have their personal “perspective” or digital workspace in which they have complete control over editing, arranging and managing their own mix of shared and private notes (short texts, graphics, Web links). There is also an official class perspective with topic headings, class minutes and agreed upon notes. Contents of the class perspective are automatically inherited (included, subject to editing) in each personal perspective. Further, it is
possible to create subgroup perspectives that reflect the work of teams or topics within the class. For instance, participants from a particular other university, from a specific academic discipline or those especially interested in one of the authors being read could set up a workspace reflecting their joint perspective on the seminar. Again, contents would be inherited from class to subgroup to selected personal perspectives. What appears in a Web browser at any given moment is a dynamic, personalized selection from the shared, interactive knowledge base (Stahl, 1999a; 2000a; Stahl & Herrmann, 1999).

WebGuide has been piloted in a middle school and a graduate school course. The results of these trials have been presented and well received at the major related conferences (see http://www.cs.colorado.edu/~gerry/webguide/publications.html). Many of the problems and limitations of previous trials will be addressed in the current semester, where a substantially revised version (WEBGUIDE 2000) will be used for the first time. This seminar is likely to produce much more meaningful data, particularly if several known problems can be addressed through the proposed project.

**Description of the Project**

The requested seed grant will allow the further development and evaluation of WEBGUIDE during the duration of the Fall 2000 seminar. By providing 50% funding for the PI, the grant will allow system development work, addition of new utilities, rapid fixes of bugs, evaluation of patterns of use, and timely assessment and reporting of successes and limitations. In addition to freeing up the PI to work on the software directly, the seed grant will let him supervise student projects related to WEBGUIDE, coordinate explicit reflection on the software by seminar participants, and prepare funding proposals to continue this research.

One major planned component of WEBGUIDE has not yet been implemented, although the technical infrastructure for it is mainly in place. That is a negotiation process whereby the group of users decide what notes should be promoted to the class perspective (or to a subgroup). A student project last semester designed this component, but it has not yet been implemented.

A major improvement to the WEBGUIDE architecture would be to convert the client/server interface to communicate using XML data rather than Java objects. This would greatly improve the ease of developing alternative interfaces to WEBGUIDE, for instance simple HTML or Perl displays and
forms. It is possible to involve a student project from an XML course this semester in working on this. The PI also has an undergraduate research apprentice who could work on this, given grant funding.

L3D is increasingly developing expertise in evaluating the success and character of online interaction. In part this is through developing methods of analyzing the structure and semantic or interaction content of discussion threads. In part it is by practicing micro-ethnographic methods of human-computer interaction analysis using digitized video recordings and computer logs. With seed grant funding, these approaches will be applied to assessing the use of WEBGUIDE and related materials in the seminar. The findings will be published and will also be used as the basis for funding proposals to the ROLE, CSS and ITR programs at NSF to continue this research. Collaboration with industry would also be appropriate and welcome at this point.

Relation to New Media Lab Mission

L3D has always pursued interaction with related industry, including NYNEX, US West, IBM, Johnson Engineering, Athenaeum International, BEA, PFU, SRA. L3D’s philosophy has been to address real-world problems and to test its ideas in real-world settings. This distinguishes it from alternative approaches oriented to abstract theory or to laboratory research.

L3D is focused on innovative research and teaching concerned with human interaction within the increasingly digital environment, whereby that environment is seen as a potential with both advantages and limitations – a future that is in the throes of being invented, and whose invention we can influence.

L3D strives to integrate teaching and research, with a strong project-based orientation in its classes and a dominant involvement of undergraduate and graduate students in its research. Vertical integration is a way of life here. Relations between L3D and industry have historically included both long-term placement of students on site and substantial visits by industry scientists on campus, in order to build lasting, meaningful relationships and deep shared understanding.

The proposed project fits nicely within the Lab for New Media’s theme of “Perception and Persona in the Digitally-Mediated Environment.” The perspectives mechanism in WEBGUIDE is designed to represent the intellectual persona of participants and to allow these persona to be perceived dynamically.
The project falls under Technology Research, involving directly the design, implementation, use and assessment of middleware in support of collaborative interaction. At the same time, in the seminar setting it is used to explore the strategic integration of face-to-face and computer-mediated interaction. Finally, within the tradition of work at L3D and by the PI, it involves the dynamic configuration of text and graphics from a shared, interactive knowledge base into a hypertext narrative structure personalized to the user’s changing interests.

References


Interoperability among Knowledge-Building Environments

Abstract:

A number of software environments have been developed to support collaborative knowledge building, typically incorporating a persistent discussion forum. Despite striking similarities and interesting differences among these community learning tools, there has been little direct interchange of ideas, designs, experiences and data among the developers. A first step toward increasing collaboration in this research community is to define a mark-up language to represent, archive and translate the data captured in these systems. This will help us to understand the design space of such knowledge building environments, to share software tools and to archive data for analysis. This project brings together representatives of research groups building related tools and evaluating the learning supported by those tools.

Principle Investigators:

PI: Gerry Stahl, Gerry.Stahl@Colorado.edu, University of Colorado (WebGuide)

Co-PI: Matthew Realff, Matthew.Realff@che.gatech.edu, Georgia Tech (CoWeb)

Co-PI: Charles Kerns, Charles.Kerns@Stanford.edu, Stanford University (Learning Lab)

Co-PI: Christopher Hoadley, tophe@unix.sri.com, SRI (Knowledge Network)

Other Executive Committee Members:

Chris Teplovs, Chris.Teplovs@utoronto.ca, U Totonto (CSILE/Knowledge Forum)
Jay Scott, jay@forum.swarthmore.edu, Swarthmore College (Math Forum)
Patricia Schank, schank@unix.sri.com, SRI (Tapped In)
Alex Cuthbert, alx@socrates.berkeley.edu, UC Berkeley (KIE/Wise)
Janet Blatter, jblatt@po-box.mcgill.ca, McGill University (LearningSpace)
Other Project Members:
Richard Wenn, rwenn@wested.org, WestEd (educational r&d)
Dan Suthers, suthers@hawaii.edu, University of Hawaii (Belvedere)
Jim Slotta, slotta@socrates.berkeley.edu, UC Berkeley (KIE/Wise)
Ken Schweller, schweller@bvu.edu, Buena Vista University (standards)
Ian McKay, help@support.maile.hawaii.edu, University of Hawaii (Maile)
Bob McLean, rmclean@oise.utoronto.ca, U Toronto (CSILE/Knowledge Forum)
Charlie Hendriksen, veritas@u.washington.edu, University of Washington (DocReview)
Mark Guzdial, guzdial@cc.gatech.edu, Georgia Institute of Technology (CoWeb)
Simon Buckingham Shum, sbs@acm.org, Open University (CSCA)
Aaron Bond, abond@interchange.ubc.ca, University of British Columbia (WebConstellations)

Objectives and Significance:

The goal is to start a collaboration among research groups and individuals who are designing, implementing, testing and evaluating innovative learning technologies that support collaborative knowledge building. A number of similar software knowledge-building environments (KBEs) have been created, generally including a discussion facility that renders student argumentation persistent. This project will define a common data mark-up format that these KBE systems can export their discussions to. This will allow interchange of data and the display of data in shared formats to facilitate comparison and research. For instance, it will facilitate the archiving of discussions from different systems in CILT's Knowledge Network.
Proposals for Research

Project and Partners:

This project will bring together researchers working within a genre of collaborative learning technology that is prominent in the larger field, but has not been identified or conceptualized as such. The effort to make the data in these KBE systems exchangeable will raise issues of software design that will lead to sharing of expertise and technical advances. This is similar to the Dexter conference that defined a software model of the major hypertext systems in 1988 and clarified directions for their future development [CACM, 37, 2]. We hope to start with KBE researchers at Toronto, Georgia Tech, Colorado, Swarthmore, Berkeley, Stanford, SRI and elsewhere. The time they contribute to this project is likely to total in excess of 400 hours during the year: drafting documents, corresponding and attending workshops. The partner groups include several people with XML experience who are willing to share knowledge of this important new technology within the KBE community. A number of groups are already working on XML representations specific to their systems, and others will soon begin to do so as a result of this project. The PI (Stahl) will be leading an interdisciplinary graduate seminar on KBEs in the Fall, with students doing research directly supportive of the proposed project. As a direct consequence of the planning of this project at CILT '99, a parallel project has been launched for research groups in the cognate field of CSCA (computer supported collaborative argumentation / design rationale), with an initial draft XML DTD already (within a week of the CILT conference) posted to a KBE for discussion. All of these activities will be represented and coordinated in the proposed project.

Expected Outcomes:

KBEs are considered important learning technologies, yet their widespread adoption remains problematic. This project will begin to bring together a community of people deeply involved in the KBE sub-field to share data, designs and experiences. Data interoperability will facilitate the development of shared tools for analyzing, visualizing and comparing student learning within various KBEs. When data is stored in an XML file, it can be interchanged between different KBE systems or different versions of the same system, archived for flexible future use and displayed on the Web with metadata search capabilities. The definition of an XML DTD for threaded discussion and related information in KBEs is explicitly viewed as just a first outcome. The KBE-ML will include a minimal model of KBE
storage, a full-featured ideal model and extensions for specific systems. If accepted, workshops proposed for CSCL '99 and elsewhere will relate this work to the broader educational issues surrounding KBEs. This project will lead to a clearer understanding of future stages of collaboration for subsequent funded projects.

**Deliverables:**

A number of participating groups have already started to work with XML representations of their own systems, so development is likely to proceed through iterations punctuated by communication and consolidation, with "standards" being repeatedly revised to support new concerns. The following milestones are targets for reaching consensus and producing semi-stable documents:

- Requirements specification for an XML DTD based on several specific KBE systems. (Month 2)
- Draft of a full-featured KBE-ML Document Type Definition for KBEs and a minimal subset to define compatibility. (Month 4)
- Development of export/import procedures between XML data and specific KBE systems. (Month 6)
- Organization of a workshop at CSCL '99 (assuming acceptance) on issues related to this project and the learning goals of KBEs. (Month 7)
- Development of Web display style sheets for the KBE-ML formatted data. (Month 9)
- Development of simple data analysis tools for the KBE-ML formatted data. (Month 11)
- Submission of funding proposals for future work. (Month 12)
- Preparation of project status updates and summary report. (Month 12)

**Organization:**

A set of four co-PIs will share primary programmatic and financial responsibility. They will decide how funds should be allocated as needs arise. Funds will be administered through the University of Colorado, but will be used to cover expenses at any participating institutions, such as the hiring of students working specifically on this project or the travel expenses.
of a participant who needs a subsidy to collaborate or attend meetings or workshops specifically as part of this project. An Executive Committee consisting of representatives of primary KBE systems will ensure the involvement of the research groups involved with those software systems. A Project Membership list of individuals will be used for the circulation of all project documents in order to build broad consensus.

**Period of Performance:**

12 months starting June 1, 1999.
Conceptual Frameworks and Computational Support for Organizational Memories and Organizational Learning

PROJECT SUMMARY

This project will investigate computer support for learning, working, and collaborating in information-intensive organizations. It will focus on communities of practice (such as local area network managers, research teams) as subgroups within and across organizations. We will work with specific communities to design, test, and reflect upon organizational memories to support organizational learning.

Organizational learning is a process by which knowledge that is created or made explicit during work on tasks is captured, structured, maintained, and evolved so it can be accessed and delivered when needed to inform future tasks. Organizational memories can facilitate organizational learning by supporting communication within communities of practice, delivering information relevant to their tasks, letting them “grow” their own information spaces, and allowing them to collaborate using the World Wide Web (WWW).

The project will work with specific communities of practice to study their actual and potential learning processes. Based on the interpretation and assessment of these observations, and theories from the research literature or from our own previous work, we will develop and articulate a new conceptual framework for computational support of organizational learning. To assess and develop this framework, organizational memories will be prototyped in collaboration with the communities and assessed in naturalistic settings.

The organizational memory software (building on emerging WWW technologies and prior research on domain-oriented design environments) will extend our currently existing prototypes with innovative mechanisms for capturing, structuring, as well as delivering information. It will incorporate computational support to reduce the burden on users as well as end-user controls to empower users to adapt the memory to rapidly evolving
needs. It will integrate the various software mechanisms into a coherent architecture and a system of meaningful user interactions for supporting effective organizational learning.

Research Issues. We will focus our research on: (1) how to capture knowledge and integrate the contexts of work; (2) how to sustain the timeliness and utility of evolving information; and (3) how to deliver relevant information actively and adaptively.

Approach. We will develop a conceptual framework for integrating working and learning in communities of practice. We will create organizational memories that include mechanisms to capture and represent task specifications, work artifacts, and group communications; facilities for practitioners to reorganize and sustain the usefulness of the memory; and techniques for access and delivery of knowledge relevant to current tasks. We will extend emerging WWW technology with structured web site interactivity, version control of evolving information, software critiquing agents, and end-user programmability.

Assessment. We will ground our designs and technical innovations in an assessment of the informational needs and organizational barriers to learning within communities of practice. We will focus our research by working specifically with communities of practice such as local-area network (LAN) designers and managers, the group of researchers working in our center, students in classes, neighborhood communities, and industrial work groups.

Expected Results. The proposed research will create (1) at the conceptual level: a unifying framework for organizational memory and organizational learning; (2) at the computational level: a generic architecture for organizational memories based on our prior domain-oriented design environments and prototypes for specific domains; (3) at the assessment level: a body of empirical results based on evaluations of the systems and the underlying theory in concrete organizational contexts.
PROJECT DESCRIPTION

Section 1. Results from Prior NSF Support

Prior NSF Awards:


Grants 2 and 4 are described below because they are most directly relevant to this research proposal.

**Grant 2. Supporting Collaborative Design with Integrated Design Environments**

**Summary of Completed Work.** This research addressed computational support for collaboration among members of design teams when direct communication among the members is impossible, impractical, or undesirable. The grant focused on the long-term, indirect communication needs of project teams rather than the needs occurring in face-to-face synchronous communication such as project meetings. Novel approaches and mechanisms were developed to capture design rationale and to associate it with the artifact to which it referred. The accumulated information was largely informal, meaning that the system was unable to perform operations on it. Techniques of incremental formalization were developed to address this problem. Incremental formalization tools were designed, implemented, and assessed that allowed system-interpretable attributes to be added to the accumulated information.

The research results of this grant were presented at major conferences in the research areas explored by the grant (e.g., CSCW, CHI, AAAI) and published in major journals in the research area (e.g., Human-Computer Interaction Journal, Applied Intelligence Journal, ACM Transactions of Information Sciences Journal). Our work has led to an increased attention in the Coordination Theory and Collaboration Technology (CTCT) and CSCW communities for long-term, indirect communication and collaboration, and design environments have been developed at numerous other places.

**Limitations Exposed by the Grant.** (1) Designers rarely capture their design rationale because it involves a lot of work and the benefits seem remote. (2) Information spaces quickly become out-of-date and disorganized. (3) Designers are not always aware that they are in need of additional information, so they make no attempt to search for information whose existence is unknown to them — hence passive information repositories are inadequate for supporting ongoing, collaborative design. These limitations motivate the proposed project emphasis on automated information capture, end-user information maintenance, and active knowledge delivery.

**Development of Human Resources.** Two members of the project are now faculty members (Scott Henninger, University of Nebraska; David Redmiles, University of California at Irvine). The graduate research assistants were offered summer fellowships at prestigious industrial research
laboratories (e.g., Xerox-Parc, NYNEX S&T, and Siemens). Two Ph.D.s were awarded (Reeves and Shipman). REUs allowed several undergraduate research assistants to be exposed and integrated into research activities.

**Five Most Important Publications:**


**Grant 4. Human-Centered Intelligent Agents Supporting Communication and Collaboration in Domain-Oriented Design Environments**

**Summary of Completed Work.** This project extended the research work and the prototype system developed in Grant 2. It explored the embedding of intelligent agents into domain-oriented design environments with the goals of reducing the cognitive load on designers through active behavior and improving the quality of the designed artifact. The project began to investigate the World Wide Web.

**Limitations Exposed by the Grant.** The project illustrated that large information repositories should not simply be built and used, but that they have to evolve by their users. This created the view that such systems should not be created by a few people doing lots of work, but should be grown by many people incrementally contributing small amounts of additional information and knowledge.
Development of Human Resources. The relatively small research grant contributed to two Ph.D.s (Christoph Thomas and Jonathan Ostwald). The graduate research assistant (Lindstaedt) worked with NYNEX on the development of the GIMMe system.

**Five Most Important Publications:**


**Relation of Prior NSF Work to the New Proposal**

In our previous grants we have explored collaboration theory and technology in the design and use of high functionality software systems to support the work of individuals and small design teams. We have developed conceptual frameworks concerning (1) multifaceted architectures for domain-oriented design environments and computational support for lifelong learning integrated with work processes [Eden et al., 1996]; (2) the maintenance and evolution of growing information bases through seeding-evolution-reseeding [Fischer et al., 1994]; (3) embedding communication in and routing work through design environments [Reeves & Shipman, 1992];
and (4) providing knowledge delivery with critiquing and other agent mechanisms [Nakakoji & Fischer, 1995].

Our research in domain-oriented design environments explored the shortcomings and limitations of generic systems and integrated different aspects of design support environments [Fischer, 1994]. Aspects investigated included active help delivery systems [Fischer et al., 1984]; critics [Fischer et al., 1991]; information filtering [Fischer & Stevens, 1991]; adaptive and adaptable systems [Fischer, 1992; Rausch, 1996]; end-user modifiability [Fischer & Girgensohn, 1990; Girgensohn, 1992; Stahl, 1993a]; and incremental formalization of large information spaces [Shipman, 1993].

The major new aspect of this proposal is to move from a primarily individual perspective (e.g., individual lifelong learning) to an organizational perspective. In our proposed project, we will develop and study a form of organizational memory based on our model of domain-oriented design environments.

In order to gain a deeper and broader understanding of the research issues associated with this shift, we organized a research symposium in May 1996 entitled “Computational Support for Continually Evolving Organizational Knowledge Bases,” which brought together a dozen of the leading researchers in organizational memory and organizational learning (for details, see: http://www.cs.colorado.edu/~ostwald/symposium/symposium.html) and we participated in a workshop at the CSCW’96 conference entitled “CSCW and Organizational Learning” [Lindstaedt, 1996b].

The proposed project builds on ideas and technologies from prior work. It expands them by focusing on organizational issues and by exploiting and redirecting the emerging WWW support mechanisms for organizational learning and organizational memories. The move to the WWW is a response to the limitations of our past closed systems, and the emphasis on practitioners sustaining information evolution is a response to the short lifetimes of our domain-oriented knowledge bases. In other, related prior NSF research (see list at the beginning of this section) we have established ongoing collaborations of our research center with community organizations, industrial partners, and interdisciplinary academic departments in Boulder as well as world-wide; the proposed project will deploy and assess our research within these organizations.
Section 2. Conceptual Framework

Our approach to organizational memories and organizational learning focuses on communities of practice as the unit of analysis, for reasons discussed in this section. We will analyze interdisciplinary sources to provide a basis for our theoretical framework, including educational theory (constructivist learning, e.g., [Harel & Papert, 1991]); design methodology (design rationale [Moran & Carroll, 1996]); cognitive psychology (distributed cognition [Norman, 1993]); social theory (activity theory [Nardi, 1993]); anthropology (situated action [Suchman, 1987]); philosophy (epistemology [Dreyfus, 1991]); sociology (communities of practice [Lave, 1988]); management science (organizational learning [Senge, 1990]); and computer science (intelligence augmentation [Bush, 1945]).

The concepts introduced in this section will be used to guide our proposed project and to assess its accomplishments. This framework suggests issues to explore, needs to support, approaches to try, and questions to evaluate. Within this context, we will design and prototype software systems to support work, learning, and collaborating in specific domains. To ground our research in the domains, we will work closely with practitioners from relevant disciplines, observing their work patterns, joining in participatory design with them, and having them try out our prototypes.

Communities of Practice

A community of practice is a group of people who share a set of activities and who interact to achieve shared objectives and to maintain their community [Lave & Wenger, 1991]. Unlike an organization, which has well-defined bureaucratic structures, a community of practice is often an informal network of people who share expertise, war stories, and practical advice [Orr, 1990]. Such communities typically form through personal ties in order to help each other keep up with new organizational or technological developments that impact their ability to get work done. These groups have a life of their own that helps them accept newcomers and survive when old-timers leave. Because of their unofficial status, communities of practice often go unrecognized and unsupported. As the role of these communities grows — particularly in information-intensive settings — it becomes increasingly important to understand them and to provide computer tools to support their functioning.

We have begun to work with local-area computer network (LAN) designers and managers at the University of Colorado in order to understand their needs for computer-supported organizational memory. This community
exists within a larger organizational structure and cuts across official boundaries based on practical needs to interact and to share information. Their information needs include technical knowledge of their work domain (e.g., what are the latest routers on the market and what are their costs, capabilities, problems, etc.), local lore (the manager of LAN x is a UNIX guru), and specific arrangements (the print server in LAN x is configured as y for reason z). The fact that most of this information is kept in the minds of individuals makes it difficult for other community members — particularly newcomers who do not yet know who has what information and have not established personal relations — to do their jobs.

We understand practice as situated activity in which practitioners pursue activity within concrete physical, technical, cultural, and interpersonal circumstances [Giddens, 1984; Lave, 1993; Suchman, 1987]. Rather than modeling practice as the execution of explicit goal-oriented procedures, we are interested in the established, generally unstated practices of a community that determine how things are done by its members — what Bourdieu calls the habitus or the tacit culture of the community [Bourdieu, 1972].

The daily practice of a community not only produces the community’s work products, it also reproduces — more or less effectively — the preconditions for the future of the community. New members learn community practices as they engage in them actively, not necessarily through didactic instruction [Schön, 1983]. As the community practice produces learning, it reproduces its own future. Because much of what needs to be passed on is never articulated explicitly, education takes place through apprenticeship relationships and training of reflective practitioners [Brown & Duguid, 1992]. This learning can be facilitated by a group memory that includes evolving artifacts of communal practice [Fischer et al., 1996b; Lindstaedt, 1996a; Ostwald, 1996].

The theory of practice addresses a number of problems that have arisen in the human-computer interaction community [Kuutti, 1996; Nardi, 1996], and that have implications for organizational memory and organizational learning. It broadens the analytic scope to take into account the social context in which people use computers [Hutchins, 1993]. The social context of a community of practice provides motivation to pass knowledge from old-timers to newcomers as everyone tries to increase their participation and reproduce the community [Lave & Wenger, 1991]. It ties working and learning together into a single framework. The introduction of new computational memories into this process will transform the social fabric, the cycles of learning, the interpersonal needs of the group [Ehn, 1989]. The design of organizational memories must take such implications into account.
Finally, the theory of practice provides a perspective on work in which sustainability means not maintaining the status quo, but rather maintaining a constant flux of new members and new knowledge. Computational environments for communities of practice must support this sustainability by allowing members to extend, update, and restructure organizational memory continuously. They must also make it easy to redefine who has access to what information in response to continual shifts in roles, assignments, and understandings. Sustainability of organizational memory means keeping it tuned to the changing needs of individuals because organizational learning takes place in parallel with the lifelong learning of community members [Senge, 1990].

Organizational Learning

Our vision of organizational learning focuses on recording knowledge gained through experience (in the short term), and actively making that knowledge available to others when it is relevant to their particular task (in the long term) [Fischer et al., 1996b]. A central component of organizational learning is a repository for storing knowledge — an organizational memory. However, the mere presence of an organizational memory system does not ensure that an organization will learn [Argyris & Schon, 1978]. Today, information is not a scarce commodity; the problem is not just to accumulate information, but to deliver the right knowledge at the right time to the right person in the right way. Organizational learning happens only when the contents of organizational memory are utilized effectively in the service of doing work [Dodgson, 1993].

Traditionally, people went to school or attended training seminars or studied books to learn facts that might be needed for later work. When working and learning are integrated in the process of organizational learning, information needed for a current task is available just-in-time [Fischer, 1991].

For sustained organizational learning, three seemingly disparate goals must be served simultaneously. Organizational memory must:

- be extended and updated as it is used to support work practices;
- be continually reorganized to integrate new information and new concerns; and
- serve work by making stored information relevant to the new task at hand.

We envision organizational learning as a continuous cycle in which organizational memory plays a pivotal role:
Proposals for Research

- Individual projects serve organizational memory by adding new knowledge that is produced in the course of doing work, such as artifacts, practices, rationale, and communication.

- Organizational memory is sustained in a useful condition through a combination of computational processes providing information (e.g., [Hill et al., 1992]) and people actively contributing [Girgensohn, 1992].

- Organizational memory serves work by providing relevant knowledge when it is needed, such as solutions to similar problems, design principles, or advice.

The intimate relation between organizational memory and work practice implies that the contents of organizational memory must be easily accessible within the context of work. Computational support for organizational learning, therefore, must tightly integrate tools for doing work with tools for accessing the contents of organizational memory.

Through everyday work, a community of practice generates knowledge that may be critical in its future [Brown, 1991]. The community’s practices are generally tacit, not written down or expressed in words [Polanyi, 1966]. Often, the only time that the knowledge exists in explicit form is when it is being actively reflected upon and used to do work [Stahl, 1993a]. By capturing this knowledge as it arises and storing it in repositories of organizational memory, a community can preserve information that is otherwise lost. Rather than building organizational memories by interviewing experts to formulate rules for expert systems, we will study the practices by which organizations do their work and communicate knowledge, and to capture the knowledge as it is articulated during work. We want to create “living” organizational memories [Terveen et al., 1995] — information spaces that are sustained and managed by the people who use them in their work, rather than by people in other parts of the organization who may have requisite technical expertise but are not intimately involved in the actual work practices [Stahl et al., 1995a; 1995b].

A principal challenge for organizational learning is to capture a significant portion of the knowledge generated by work done within a community. Experience with organizational memories and collaborative work has exposed two barriers to capturing information:

- Individuals must perceive a large enough direct benefit in contributing to organizational memory to outweigh the effort [Grudin, 1992].
The effort required to contribute to organizational memory must be minimal so it will not interfere with getting the real work done [Carroll & Rosson, 1987].

The consequence of these barriers means that processes of information capture, structuring, and delivery must be computationally supported as much as possible or they will simply not get done.

**Organizational Memory**

Organizational memories are information systems that are used to record knowledge for the purpose of making this knowledge useful to individuals and projects throughout the community of practice and into the future [Ackerman, 1994]. Ideally, an organizational memory allows individuals within the community to benefit from the experiences and insights of others, by actively informing work practices at the point when the information is actually needed [Fischer et al., 1996a]. That is, an organizational memory should not be simply a passive repository of information, but an interactive medium within which collaborative work can actually be conducted and through which communication about the work can take place and be situated.

It is often assumed that the Internet solves the problem of organizational memory. While the World Wide Web (WWW, web) on the Internet functions primarily as a broadcast medium and therefore lacks the interactivity needed, intranet structures can indeed be designed to implement organizational memories. An intranet is a small version of the web, in which access is restricted to a particular community. It uses the same technology standards (e.g., TCP/IP, HTTP) as the global web. Generally, intranet information is stored in a database rather than in fixed HTML documents, so it can be displayed dynamically to use the latest information and to respond to unique queries. Intranets are rapidly replacing traditional client/server systems as the preferred technology for computer-based organizational memories. Intranets make more flexible organizational memories because users can access them with a web browser on any computer and because the computation of the client display logic, the organization’s business rules, and the database query logic can execute on different computers.

All the major software companies are rushing to support the building of intranets. Microsoft’s Office 97 applications, for instance, can publish web documents directly. Database environments are beginning to support live data editing through forms on the web (using ODBC and JDBC database connection standards). Special environments such as Tango allow a
developer to design web data entry forms quickly using visual drag-and-drop tools. Finally, an extraordinary wave of sophisticated development environments incorporating scripting languages are appearing (at least in beta or vapor ware): Borlandís IntraBuilder, IBM/Lotusís Notes/Domino, Netscapeís LiveWire, Microsoftís FrontPage, Oracleís InterOffice, Novellís GroupWise, PowerBuilder, Cold Fusion, SuperNova, etc.

Intranet technology seems to offer a promising approach and substrate for building organizational memories. However, these environments do not by themselves suggest how to integrate work and learning, how to capture new information, how to support information evolution, how to deliver relevant knowledge, or how to computationally support these processes under user control. Yet, that is precisely what is needed. We maintain that systems to support organizational learning should take an analogous approach to our domain-oriented design environment support for informing collaborative design work. We propose to explore organizational memory that does this, using commercially available intranet technology as an enabling technology.

Section 3. A Scenario of Organizational Learning Using Organizational Memory

To address the issues reviewed in the previous section, we propose to prototype an organizational memory system named WebNet that explores these issues within concrete work contexts. One community of practice with whom we plan to collaborate in designing and assessing WebNet is local-area network (LAN) designers and managers at the University of Colorado. Following is a vision of how WebNet might be used by this community. The scenario illustrates how WebNet integrates working, learning, and collaborating. The purpose of the scenario is to present concrete examples of the kinds of information and mechanisms that WebNet will include, as a background to the discussion of computational support in the following section.

Kay is a geography graduate student who works part-time for network services. Kay logs into WebNet through her web browser, and WebNet responds by displaying Kay's WebNet home page. Kay had designed this page to include information sources she needs to check regularly; it delivers information that is related to her LAN and to her job responsibilities. Kay's WebNet home page contains a message list (with email and comments directed to Kay from co-workers and clients), a to-do list for tracking her
current projects, and a community-wide task list of jobs that need to be done.

**Integration of the Work Situation.** Kay notices that she has a message from Ray, her supervisor, suggesting a new task for her. Kay selects the Geology job from the task-list and WebNet displays a task specification page (see Figure 1-A). The task specification says that a new Windows NT Server and three Macintosh PowerPC workstations are to be connected to the Geology Network in room 214. Kay's task is to prepare a logical design, parts list, and price breakout for the new installation. The task specification also provides a budget and contacts within the Geology Department. Kay clicks on "reserve task" to inform her co-workers and WebNet that she will take care of the task.

When Kay clicks on "Geology Net" in the task specification, WebNet displays a logical map of the current Geology LAN in the knowledge-based construction tool for LANs. The construction tool provides a work area, a tool bar, and a palette of network design elements that can be selected with the mouse and placed in the work area (see Figure 1-B).

Kay begins to plan the installation of new equipment by adding the purchased equipment to the existing logical network using the construction tool. She selects the Macintosh icon from the palette and places three workstations into room 214. Then she selects a Windows NT icon and places it. Finally, Kay connects the new equipment by dragging the cable to reach from the existing network to each of the new machines.

**Information Delivery.** When Kay has connected the machines to the network, WebNet beeps and places a blinking router icon at the junction between the existing network and the portion that Kay has added. A critic message appears in WebNet’s lower pane, indicating that the configuration she has specified requires a router. Kay knows what a router basically does and why a router is needed in this configuration. However, she doesn't know what specific router is needed or how much the needed router should cost. Kay selects the link to “router” in the critic message, and WebNet brings up a new page containing information about routers (see Figure 1-C). The router information page contains a short description of routers from the WebNet glossary, a collection of definitions for common networking terms.

Kay finds that this definition is also too general so she decides to check out some displayed bookmarks. Bookmarks consists of a catalog of URLs that previous WebNet users had found helpful and had added. WebNet has displayed the bookmarks that are relevant to the current design. To Kay's disappointment, the bookmarks point to router manufacturers’ pages, which
contain detailed specifications about the routers, but not the type of information that Kay needs.

Kay decides to search WebNet's information space. WebNet supplies a default query based on the current LAN design context: "list all information about routers". Kay can use this default to search WebNet, or she can modify the default query by simply typing in more words to the query box. More sophisticated searches may be performed by selecting the "more choices" button, which brings up a query window containing an interface for constructing queries involving particular information sources within WebNet, author, dates, and specific networks, in addition to the search string.

Kay begins her search by selecting the "Search Now" button. WebNet displays links to many pieces of information, ordered by their relevance to the query string. Overwhelmed by the amount of information, Kay decides to refine the query. She selects "more choices" and restricts her search to email written in the past six months and modifies the query to "list emails about routers for small networks" (see Figure 1-D).
Proposals for Research

B.
C.

D.
This query returns just twelve email messages. One describes how Pat used a PC as a router in a small LAN. Pat's email indicates that routing in software can be cheaper and more flexible than through a hardware router, although there is a performance penalty.

**Sustaining the Organizational Memory.** Kay decides that Pat’s solution may also work for the Geology Network. She adds information about her solution to WebNet’s glossary, making it available to other members of the network design community. She includes a link to Pat's email message, and also a link to her design, to connect these related pieces of information. Now other designers in Kay’s community will benefit from the knowledge Kay uncovered through her work.

Kay returns to the task description page and adds a status report describing her proposed design. She tags the status report to be sent to Ray, to the Geology contact person, and to Pat, asking for feedback on her decision.

---

**Section 4. Computational Support**

The conceptual framework presented in Section 2 implies that organizational memory systems supporting organizational learning must be tightly integrated with tools for doing work in order to capture new knowledge, to allow the community to sustain it, and to actively deliver information when needed. Only if the organizational memory includes representations of the work context can it decide what information is relevant to the current task. This project will explore mechanisms for the software to make such determinations. The scenario showed a simple example of one person interacting with such an organizational memory.

This section presents our technical approach. Our approach extends prior work by us and by others; it also takes advantage of emerging intranet technologies. The significance of this project is to integrate the techniques in a theoretically motivated way and to assess how well they can address the practical issues confronting communities of practice in the information age. After reviewing our prior work and the work of others, this section will discuss some mechanisms for addressing our core research issues:

1. How to capture knowledge and integrate the contexts of work.
2. How to sustain the timeliness and utility of evolving information.

3. How to deliver relevant information actively and adaptively.

**Relation of Our Prior Work to Proposal.** We have created and assessed design environments for the following domains: kitchen design [Nakakoji, 1993], programming [Mastaglio, 1991], user interface design [Lemke, 1989], voice dialog design [Sumner, 1995], simulation design [Repenning, 1993], lunar habitat design [Stahl, 1993a], service provisioning [Ostwald, 1996], and LAN design [Reeves, 1993; Shipman, 1993; Sullivan, 1994]. In particular, we investigated how such systems could support the location [Henninger, 1993], comprehension [Redmiles, 1992], modification [Girgensohn, 1992], querying [Fischer & Nieper-Lemke, 1989], filtering [Stevens, 1993], and sharing [Rausch, 1996] of information in large information spaces of domain knowledge.

This prior work has contributed to a prototypical architecture for domain-oriented design environments, which integrates working and learning with components for: (a) construction of the design artifact, (b) a knowledge base of design rationale and artifact designs, and (c) computational critics that actively deliver relevant knowledge. The proposed project will generalize from this prototypical architecture to one having the following general functions: (a) representations of the work/collaboration context, (b) a rich, sustainable information space, and (c) mechanisms to map from the work context to relevant information for delivery.

Just as our work on design environments involved the interplay of multiple software components (e.g., construction, simulation, specification, gallery, catalog, rationale, critiquing, etc.) to deliver relevant design rationale, the proposed project will investigate mechanisms that deliver timely knowledge to practitioners by retrieving items from the information space that are related to the current work context. We will discuss our approach to implementing these mechanisms in the remainder of this section, after we relate our approach to that of others.

**CSCW and Distributed Artificial Intelligence.** Our approach distributes work and control between human practitioners and computational agents embedded in organizational memory. Our paradigm shares a large number of research issues with two related areas: (1) Computer Supported Cooperative Work (CSCW) [Greif, 1988], which emphasizes communication and collaboration among humans mediated by computer; and (2) Distributed Artificial Intelligence (DAI) [Bond & Gasser, 1988], which emphasizes communication and collaboration among computational agents. In order to enrich CSCW environments with computational agents,
Semi-formal Systems. Our approach to formalizing information in the organizational memory attempts to avoid the need for complete formalization without placing an unmanageable burden on the people who use the system. By representing the contexts of work, we establish a shared understanding of that context by the system and its users. By combining automatic capture of information, incremental formalization of stored knowledge, and end-user control over structure, we try to facilitate a workable balance. Related work on semiformal systems indicates that formalization need not be complete to be useful in aiding communication and collaboration [Bobrow, 1991; Malone et al., 1988; Malone et al., 1992; Winograd, 1988].

Workflow Systems. Our approach to organizational memory can be contrasted with workflow systems [Ellis, 1991] and other systems that are established by an organization to structure and regulate work processes. Workflow systems may be appropriate to coordinate regular and predictable interactions among different work groups, but they are not appropriate to support the situated and often ad hoc work within communities of practices, where innovation and change are ubiquitous. Similarly, systems to support ISO 9000 often try to incorporate organizational memory about work procedures in client/server or intranet systems. However, they center on a hierarchy of documents and fail to capture the implicit practices, tacit background knowledge, and changing circumstances that are critical to organizational practice. Organizational memory systems for communities of practice should empower users to cope with vague problems and unexpected breakdowns, and to share innovative solutions and work practices with their peers.

Design Rationale. Our approach also contrasts with design rationale schemes such as gIBIS [Conklin & Begeman, 1988] that require designers to interrupt work to articulate justifications for their design moves [Fischer et al., 1996a; Reeves & Shipman, 1992]. Designers are often unwilling to invest the extra effort to provide rationale [Grudin, 1994]. Similarly, most web-based group memory systems — such as threaded conversations and Frequently Asked Questions (FAQs) — are divorced from work contexts, so they can neither capture knowledge as it is articulated nor target retrieval to work states. Moreover, like many other information systems, these are
impoverished in that they cannot contain work artifacts themselves, but only discussions about artifacts.

We postulate that organizational memories need to incorporate tools for working and communicating within the system. This is something that workflow, ISO 9000, design rationale, and similar support systems fail to do. By including software components for design, analysis, communication, etc. in which community members can carry out some of their work and through which they can collaborate with each other, organizational memories can address their central tasks: to capture, sustain, and deliver information.

1. How to capture knowledge and integrate the contexts of work

We will combine several mechanisms for embedding work and communication in a computational information system that we implemented and assessed in our previous NSF grants. In Janus [Fischer et al., 1989] and similar design environments, the construction of a design artifact takes place in a construction component that uses a palette of domain items so that the software can track the semantics of the design. In the Remote Exploratorium [Ambach et al., 1995], the domain items in this palette can be exchanged within a virtual community through a web page within the system. In the Indy system for LAN design [Reeves & Shipman, 1992], post-it notes and other annotations can also be embedded in the construction area. The Kid system [Fischer & Nakakoji, 1991] incorporated a specification component to capture and represent design goals. EVA [Ostwald, 1995] routed design ideas through a shared computational repository. In GIMMe [Lindstaedt, 1996a] email is sent through and archived in a group memory. These mechanisms can all be used in organizational memories. The scenario illustrated several. Kay worked on designing the extended LAN within a construction/simulation component and she found critical information in an email component based on GIMMe.

Our organizational memory systems will generalize the notion of representing the contexts of design. In addition to representing the layout of an artifact or its specification criteria, a system can, for instance, represent the people involved — either as individuals or as workers in certain organizational roles. The Hermes design environment [Stahl, 1993a] explored a perspectives mechanism that tagged versions of information as belonging to different perspectives: different system users chose to retrieve information according to their profession (e.g., plumbing or electrical); domain (residential, commercial, industrial habitats); or organizational role (designer, supervisor, manager). Situating knowledge delivery within Janus,
Kid, or Hermes-type contexts — constructions, specifications, perspectives — can facilitate the selection of relevant information.

In the context of LAN design and management, WebNet representations of the problem context will include: physical layout of equipment; logical layout of functional components; simulation of major network traffic sources and routers; performance specifications; professional perspectives; organizational business rules; problem reports; and email discussions. Each context representation will require its own user interface to allow people to modify the characteristics and effects of the representation as well as to instantiate representations of specific tasks. Each representation will affect the selectivity of the system’s knowledge delivery.

We will use mechanisms for communication capture such as those we used in the GIMMe email archive. GIMMe [Lindstaedt, 1996a] works this way: a community establishes an email alias for all communication of general interest to the group. In addition to members getting the email, it is also sent to a group memory archive. Here it is indexed for full-text search (using latent semantic indexing, described below) and made available for searching and browsing by community members. Members can also reorganize the mail by categories. Not only can members stop reading their daily group email and periodically scan GIMMe by categories of interest, but new members can learn the group’s history, and all members can retrieve prior discussions and decisions. GIMMeis functionality will be incorporated in WebNet, where it will be enhanced with tools to sustain its evolution and to actively deliver relevant contents based on work contexts.

2. How to sustain the timeliness and utility of evolving information

The approach to sustaining information is based on an extension of our model of system evolution (iSERi). We want to empower practitioners to evolve their own information spaces in a sustainable way. This requires making the web interactive (with iDynaSitesi) so information can be changed as it is used. It also requires structuring mechanisms (such as iperspectivesi) for organizing changing information.

Sustainable Evolution. In evaluating our domain-oriented design environments, we observed that the information stored in the knowledge bases soon became obsolete, as did the system functionality itself. Our seeding, evolutionary growth, and reseeding (SER) model [Fischer et al., 1994] is an attempt to see how community members can evolve their information systems [Henderson & Kyng, 1991]. The model distinguishes
three categories of professionals involved in creating, maintaining, and using an organizational memory:

- **Substrate producers.** These are the people who create the underlying technology. For our project, these are the producers of intranet development environments and other substrates.

- **Memory designers.** These are the people who design and implement an organizational memory. For our project, these are members of our research group.

- **Practitioners.** These are the people who use the organizational memory in their work practices. For our project, these are the communities of practice that collaborate with us and try out our prototypes.

The SER model consists of the following three processes:

- **Seeding.** In the seeding process, memory designers and practitioners work together to instantiate an organizational memory seeded with domain knowledge and local information.

- **Evolutionary Growth.** In the evolutionary growth process practitioners add information to the seed as they use it to do work. Work artifacts and communications accumulate in the organizational memory, resulting in growth of memory contents. In addition, new work produces needs for new system functionality and structures.

- **Reseeding.** In the reseeding process, memory designers and practitioners reorganize and reformulate information so it can be reused to support continuing tasks.

In the proposed project we want to investigate the possibility of going beyond our prior reseeding model by providing mechanisms for communities of practice to sustain the growth of their organizational memories without a distinct reseeding phase. Organizational memories should be able to evolve in symbiosis with their communities of practice like biological species evolve with their environments — with no interventions needed from outsiders. Incremental formalization techniques [Shipman, 1993] can be used to automatically add computationally interpretable attributes to information that has accumulated during the evolutionary growth. Formalization of information increases the system’s ability to structure and retrieve the information, and thereby generalizes the information content beyond the specific context in which it was originally added. Empirical evidence shows that, within communities of computer users, technically proficient ilocal developersí emerge who are willing and able to perform many system modifications [Nardi, 1993]. WebNet will
include mechanisms for communities — especially their power-users — to use to sustain the usefulness of the organizational memory continuously, thus reducing the need for a separate, disruptive reseeding phase that requires the memory designers to return. We have begun to explore this possibility with GIMMe, which allows members to reorganize as well as search and browse its email repository. We will make use of open industry standards so we can take advantage of future technological advances to increase substrate functionality, too, with minimal disruption.

**Interactive Web Sites.** Technologies for intranets are proliferating. However, these commercial products are very generic enabling technologies. They provide tools for building organizational memories but do not by themselves solve the complex issues of capturing, structuring, and delivering information.

![Figure 2. The web as (a) broadcast medium and the web as (b) interactive organizational memory.](image)

WebNet will address current limitations of the web for supporting the kind of interactions required for communities of practice to use and sustain their organizational memories. As typically used, the web is a broadcast medium; people can find some information there, but they cannot easily modify, update, restructure, or contribute to the information source (see Figure 2). This model of the web suffers from one-way communication, poor coverage, poor reliability, low relevance, static format, and rapid obsolescence. These are serious problems for organizational memory. We propose an interactive model of the web, designed to solve each of these problems [Ambach et al., 1997; Fischer & Thomas, 1997; Stahl, 1997].

WebNet will use what we call “DynaSites” to create a rich, interconnected, searchable, and browsable organizational memory that is easily updated and
annotated. DynaSites are dynamic web sites. The HTML pages viewed in standard web browsers are created on-the-fly by a commercial intranet builder from data stored in a relational database. Wherever appropriate, the pages include forms for viewers to make changes to the content, subject to a system of permissions that can be configured for each DynaSite. These forms update the underlying database (and thereby the content of future web pages) without requiring any database or web technical knowledge by the users. The structure and other characteristics of a DynaSite can be set up in advance by memory designers or defined and modified by practitioners with appropriate authorizations.

The specific content of DynaSites in a given system will depend upon the domain and the community being supported. Initial DynaSite structures will arise through a participatory design process in which we work closely with community members. Based on our initial explorations of DynaSites for local area network managers, their information space will include a glossary of technical terms and local terminology; a repository of all group email; the current queue of trouble reports and its history; a table of configuration data for each host on a LAN; a diary of changes made to each machine; an online manual of approved methods and procedures; LAN design rationale; a FAQ (Frequently Asked Questions about local LANs); threaded discussions among group members; and a directory of external pages of equipment vendors and other external web sites of interest.

**Perspectives.** In prior work we have explored a “perspectives” mechanism [Stahl, 1993b; 1997] that will be adapted to DynaSites. Perspectives are important for sustaining evolution in collaborative information spaces [Boland et al., 1992]. They allow different changes to the information to be maintained simultaneously in different perspectives. This way, people can make successive changes to the content or organization of information without negating the effects of previous changes. For instance, if one design group has completed an artifact that satisfies all relevant critic rules and saves the artifact in their group perspective, then later changes to the critic rules by another group will not affect the subsequent evaluation of the artifact within its original perspective. In such cases, perspectives provide a versioning system for organizing and sustaining a memory system that evolves over time.

Perspectives can also be used to make knowledge delivery relevant to subgroups of a community. For example, everyone who maintains a given LAN or set of LANs can use the same perspective; there can be a perspective for Macs, one for PCs, and one for UNIX management; perspectives can correspond to the chart of organization, with supervisors
having more modification permissions and oversight over certain DynaSites. Perspectives inherit from each other hierarchically [Bobrow & Goldstein, 1980], determining what information is retrieved, how it is displayed, and what modification permissions are granted.

3. How to deliver relevant information actively and adaptively

The standard mechanisms for retrieving information from the Internet or from intranets yield frustrating results [Berghel, 1997]. Indexers and search engines such as Yahoo and Alta Vista work best when information is structured — but the web is not. Typical first searches return hundreds of thousands of hits, with follow up queries returning either still unmanageable thousands or none at all. Attempts to ipull down information of interest automatically using software agents (bots) is not yet practical — there are still too many unresolved issues involving how to specify relevance through end-user programming or otherwise. While there is much current work on software agents as a means to aid users in locating information in large information spaces (e.g., [Fischer & Thomas, 1997; Maes, 1994]), most of this work either relies on the user to explicitly formulate a query or relies on an implicit user model. iSolicited push[i Wired, 1997] through subscription to specialized or reliable information services is likely to be a popular solution for receiving domain news, but it does not meet the needs of just-in-time learning. Once more, none of the generic solutions integrate information delivery with work.

In our prior work, we have addressed the retrieval problem with query by reformulation [Fischer & Nieper-Lemke, 1989], filtering [Fischer & Stevens, 1991], and critics [Fischer et al., 1991]. In organizational memories, we want to empower practitioners to take maximum advantage of shared knowledge. However, we do not want users to have to formulate database queries as such — that requires professional training and knowledge of data storage structures. We are interested in providing as much software support as possible in formulating initial queries, letting users select from catalogs of queries, and helping people to reformulate queries at a level of abstraction corresponding to how they think about their work tasks. This means integrating the information delivery process with the work context with mechanisms like critics. Another mechanism for doing this is suggested by latent semantic indexing (LSI). Finally, we propose to develop an end-user scripting language for practitioners to reformulate queries.

Critics. Our domain-oriented design environments used the context of constructed artifacts, specified design goals [Nakakoji & Fischer, 1995], and
selected perspectives [Stahl, 1993a] to guide retrieval [Fischer et al., 1993]. Computational critics in these systems are agents that monitor the changing work context and identify potential information needs; when such a situation is identified the critic offers to deliver relevant information. Specifications can be used to select different sets of critics, and perspectives can reinterpret the behavior of the critics [Fischer et al., 1993]. Critics remain an important mechanism for organizational memories, but we want to find additional mechanisms to map from representations of the work context to relevant information.

**LSI.** To combat the brittleness of keywords in searches, we use latent semantic indexing (LSI) [Dumais et al., 1988; Landauer & Dumais, 1997]. We have experience using this with GIMMe, where it provides the primary access to archived email. LSI works by building a multidimensional scaling space through a statistical analysis of all the vocabulary in a textual corpus such as an organizational memory. Using this, LSI can locate items that are closely related to a given word or a longer phrase; it is not restricted to items that contain the exact keyword. LSI nicely augments the use of an embedded work context to help locate relevant information. For example, a textual problem report or a specification document can be used directly as an LSI query to retrieve stored documents that are semantically related (i.e., that deal with the same problems or with related machines and people). In this way, a document in the work area, such as a task description or a problem report, can be used by LSI to find other documents (emails, procedure manuals, previous problem reports) that are related and could prove helpful.

**Scripting Language.** Because information-intensive work is creative and communities of practice are dynamic, the retrieval of needed information must be under the control of the users. An organizational memory should allow practitioners to modify the information retrieval processes themselves. We will include an end-user scripting language to allow non-programmers to formulate and modify queries. We developed similar scripting languages in our Hermes [Stahl, 1993a; Stahl et al., 1992] and Agentsheets [Repenning, 1995] design environments. The syntax and vocabulary of the language will reflect the structures of the DynaSites database schema and representations of the work contexts, but a drag-and-drop interface to the language will shelter the user from worrying about these matters. In the scenario, for instance, Kay formulated the query, "list emails about routers for small networks."
Section 5. Assessment in Practice

Our project approach incorporates ongoing assessment of our conceptual framework and computational mechanisms. The framework suggests important requirements and mechanisms; our success in designing the mechanisms and the results of assessing them in use will reflect back upon the theory, highlighting important issues for organizational memories and the communities that use them.

We will assess our conceptual frameworks and prototypes in a variety of settings for organizational learning, such as those discussed below. In each of these settings, efforts to enact organizational learning will focus on reconceptualizing the use of technology for organizational learning, rather than simply “gift-wrapping” traditional frameworks with new technologies.

LAN Design Community. The domain of LAN design and management is appropriate because work is done by a community of practice; LANs are not designed once and for all but evolve over time; LAN design relies upon an enormous and rapidly changing information base; and LAN managers do much of their work on computers. Within this setting we will assess the integration of working and learning, new forms of collaboration enabled by our systems, and the ability of the community to sustain their computational environment over time.

Boulder County Healthy Communities Initiative. BCHCI is a community-based effort (of approximately 500 citizens coming from different backgrounds) to identify major trends and implement positive change on issues that affect the long-term health, quality of life, and sustainability of Boulder County. The concerns of BCHCI are (1) to engage citizens as self-directed learners who understand sustainability and can actively participate in design solutions, and (2) to turn BCHCI into a learning community that benefits from citizen input. Our relationship with BCHCI provides a unique opportunity to establish an organizational memory and study organizational learning within a community setting.

NYNEX University. We will build upon our ten-year relationship with NYNEX to reconceptualize organizational learning in industrial settings. NYNEX is the regional telephone company for the New York/New England area, with about 50,000 employees. NYNEX has made an unprecedented commitment to lifelong education of its front-line workers by establishing NYNEX University campuses throughout its operating regions. GIMMe technology is used as part of this effort, aimed to train workers to keep up with the rapid changes in their field through (1) a deeper understanding of emerging technologies, (2) competence with computational tools for finding
and communicating new knowledge, and (3) a new emphasis on peer-to-peer learning in the workplace.

**L3D Center** (including the proposed project). Our research center aims to develop computational support and conventions of use that enable us to be a learning research community. We will create organizational memories for our center, as well as for the proposed project. This self-application of our theories will give us first-hand experience with the strengths and limitations of our conceptual framework and technology.

**University of Colorado Courses.** University courses have traditionally been based on instructionist educational strategies, emphasizing fixed curricula, memorization, and decontextualized learning. The proposed project will continue our standing commitment to exploring new models of education that emphasize peer-to-peer learning through projects and discussion-oriented classrooms. We will use project prototypes in our own classrooms, where students will explore and reflect upon innovative applications of organizational memory.

**Section 6. Work Plan**

**Year 1.** Our initial focus will be on careful analysis of the current practices of the LAN design community at the University of Colorado and our own research group. We will work with an anthropologist to understand the existing practices of communities we collaborate with. Our goal will be to extend the unit of analysis from an individual working and learning to an organizational focus. System-building efforts in the first year will focus on the implementation of a core WebNet system. We will employ available technologies and our prior system mechanisms and extend them as needed.

**Year 2.** In the second year our emphasis will be on envisioning and enabling new ways of working, learning, and collaborating. We will work closely with several communities to create organizational memory seeds. The seeds will define initial community-specific organizational memories. The seeding process will be grounded by the creation and collaborative assessment of prototypes, with communication about the prototypes captured within the organizational memories. This approach will interleave system-building and assessment, and capture a history of the seeding process that will serve project assessment as well as the ongoing evolution of the systems. We will embed logging mechanisms in the organizational memories to facilitate tracking of the evolution of both information content
and structure. In this year we will extend our system functionality with perspectives and a scripting language.

**Year 3.** The final year of the project will have two primary foci: (1) the use and sustainability of organizational memories by the communities of practice, and (2) an integrative framework for organizational learning in a variety of settings. Assessment of organizational learning in communities will be both quantitative and qualitative. Logs of information use and evolution will provide data about the mechanisms most used and about the dynamics of the organizational memory. By analyzing the usage logs in conjunction with user interviews, we will assess how well the mechanisms and systems supported the sustainability of these information spaces as useful sources of organizational memory under changing conditions. Our assessment of organizational learning in the research settings will lead to an integrated conceptual framework for organizational memories and a generic architecture of computational support for organizational learning.

---

**References**


Specification Component, Ph.D. Dissertation, Department of Computer Science, University of Colorado at Boulder.


Allowing Learners to be Articulate: Incorporating Automated Text Evaluation into Collaborative Software Environments

Proposal to the McDonnell Foundation

A joint project of the Institute for Cognitive Science and the Center for LifeLong Learning and Design at the University of Colorado

1. Abstract

We have been developing software environments that allow teachers or students to build educational simulations to foster collaborative learning. In particular, our WebQuest adventure games motivate players to explore subject matter topics on the World Wide Web as part of classroom research projects. Our goal is to explore how “edutainment” software (like Where in the World is Carmen Sandiego?) can support the construction of personal knowledge and the articulate self-expression of learners. In this effort, we have confronted a problem that is quite pervasive in educational software: the challenges posed to game players by games like WebQuest are currently restricted to questions having well-defined factual answers that can be checked by the software. In order to promote and evaluate the construction of deeper knowledge the software needs to be able to make computational judgments about the content of unrestricted essays that the students write.

A new mathematical technique being developed as part of a cognitive theory of text comprehension/latent semantic analysis, or LSA/promises to provide the necessary computational ability. LSA computes the semantic relations within a corpus of literature on a given subject matter and then uses this information to judge the semantic similarities among submitted written responses. Although LSA has been found to be almost as reliable as human readers in several laboratory tests, it has yet to be applied in classroom settings. The proposed project will incorporate LSA in a variety of ways within our educational software in order to explore a range of theoretical issues related to how computer-based media can help students learn.
We will develop several of our current software prototypes (WebQuest, Remote Explorium, Teacher’s Curriculum Assistant) further by extending them with LSA mechanisms and by working with teachers and students in the classroom. Development will be guided by cognitive theory concerning text comprehension, research techniques for educational software, and evaluation of various applications of our software in educational practice. The software will be extended to allow students to design and create their own games for fellow students to play. Both questions and answers will be in text format, evaluated automatically by the software using LSA. Classes can select themes, create multiple games incorporating summaries of group knowledge, critique the games, and share the games with other schools over the Internet. Ultimately, LSA can be used to match the most appropriate versions of games or information sources on Web sites to different classrooms or to individual students by evaluating the students’ written products and comparing them to alternative sources of background information.

The project goal is to explore computer-based tools for supporting the collaborative construction of knowledge in classrooms and the articulate self-expression of individual learners without over-burdening teachers. Automated text evaluation mechanisms will be investigated to allow fact-centered questions to be replaced with open-ended, question-answer interactions, without requiring continuous teacher intervention. More generally, the project will address how software environments can help students to learn in an information-intensive, technologically mediated world by matching individual competencies to appropriate resources.

2. Instructional Problem

The Center for LifeLong Learning and Design at the University of Colorado has been working with classrooms and teachers in the Boulder Valley School District to conduct research in educational software. Specifically, our WebQuest software presents students with an adventure game that teaches students research skills involving the Internet. (See Figure 1.) Each time a student confronts an obstacle in the game, the student must answer questions using information found on the World Wide Web (WWW or Web).
Figure 1. A scroll in WebQuest and a WWW page it suggests for finding the answer. The scroll is presented when a WebQuest player encounters a challenge. The player must conduct research to answer the question. The scroll suggests Web sites or search engines to guide the player to relevant pages on the Web.

Students are enthusiastic about playing the game and surfing the Web. Although WebQuest was just recognized as the “best innovative application of the WWW for education” at the international WWW5 conference in Paris, we think we can make it into a much more effective classroom tool. We recognize several major pedagogical weaknesses to our current approach, based on constructivist theories of learning. These weaknesses are endemic to the computer game approach to education, in which one tries to embed learning opportunities within a motivational game context:

- The questions posed require multiple-choice or keyword answers, not the articulation of deeper reflection.
The investigation of information is guided by an externally imposed game framework, rather than being student-centered.

The acquired knowledge is not tuned to the background knowledge and capabilities of the student.

The learning process is not social and interactional.

As a first step in overcoming these weaknesses, we have begun to experiment with having students actually author (i.e., design and program) adventure games for their classmates to play. This makes for a much more intense learning-by-teaching experience; it opens up exciting new possibilities for interactions in the classroom. However, the bottleneck of multiple-choice or keyword answers remains. A student authoring a game must reduce any knowledge about a topic to a few atomic facts which students playing the game have to match literally. We want to allow learners to be more articulate than this.

Multiple-choice questions and keyword answers have always been resorted to in education for pragmatic reasons. Teachers simply do not have the time to read and understand answers to open-ended questions for every test and quiz. Multiple-choice questions have been used for standardized tests because of technical limitations to machine processing of answer sheets. We know how stultifying this restriction to keyword answers has been. It forms a major barrier to moving classroom emphases from the memorization of atomic facts and isolated terms to the construction of deeper understanding and fuller self-expression.

The constructivist alternative to multiple choice questions has proven untenable to date because of the burden it places on teachers. Within the context of an NSF-funded research project focused on learning-on-demand (student-centered and task-centered) we found that self-directed, authentic learning activities require substantially more teacher resources than are normally available in K-12 or university classrooms. Teachers must evaluate written reports and portfolios on topics that may be relatively new to the teachers themselves. To be most effective, feedback in response to student attempts at articulating their growing knowledge must be timely. In the context of educational games, the situation is even more extreme: evaluation of answers must be immediate to avoid interruption of the motivational game context.

If educational software could adequately process unrestricted text, then it could provide a medium for students to construct and communicate higher-order understandings of subject matter without placing an impossible burden on teachers. For instance, if WebQuest could automatically evaluate
unrestricted text, then authors of new games could define obstacle problems using short essays, and students playing the game could enter brief texts that would be compared with the problem essay. In this way, everyone could express their own understanding in their own terms.

The ramifications of evaluating unrestricted text by educational software are far-reaching. Ultimately, this capability would allow textual presentations of topics to be selected based upon students’ background knowledge. For instance, an individual student or a classroom of students could be evaluated by software that analyzes their sample writings. When the software then presents WWW sites for the student to explore, it could select sites whose text is at an appropriate reading level. As the use of such software becomes prevalent, WWW sites, WebQuest games, and other educational resources could be structured to provide versions of texts at different reading levels. In the “articulate classroom” that we envision, students would express their ideas in writing, producing portfolios of text that the software could evaluate to form a model of the students’ levels of understanding. This would provide a valuable tool for the teacher to use in guiding students.

The fact that software like WebQuest is currently restricted to multiple-choice questions illustrates a significant and wide-spread problem in education: how to evaluate, score, classify, and otherwise process unrestricted text automatically, without laborious efforts by highly qualified but over-burdened professionals, such as teachers. Adding a free text capability to WebQuest could increase the educational value of the software, in that information of greater complexity could be searched for, and the students’ answers would not have to be as narrowly constrained.

We believe that full natural language understanding by computers is not necessary to remove the bottleneck. Certain computationally feasible analyses of text may be sufficient to meet the needs of software like WebQuest for processing essay answers. The proposed research would permit us to explore this possibility, further developing a promising text analysis technique and extending our educational software to overcome its current weaknesses. Moreover, the project would allow us to test and refine our laboratory-based theories of text comprehension within the context of classroom practice.

Specifically, we propose to investigate a new technique of text evaluation known as latent semantic analysis (LSA). We anticipate that LSA can provide a fully automatic computer technique that allows assessing the content of a text by comparing it with other texts, such as books, articles, essays written by students, single sentences or phrases, even single words. The technique has its limitations and is still being developed. Furthermore,
we have only begun to explore its implications, both for psychological theories of meaning and for educational applications. Nevertheless, our work has progressed enough to show that further research along these lines is worthwhile and, indeed, highly promising.

The general cognitive issue that we want to focus on with the proposed project is the question of what it means to acquire subject matter knowledge using tomorrow’s technologies of large information bases and efficient search methods. The ability of students to benefit from external information sources both relies upon a level of internally assimilated background knowledge and simultaneously transforms the motivations for acquiring and internalizing such knowledge. What content do students have to know for successful searching? Will they learn if they know that they can always easily find answers by searching? How do these factors combine to produce intellectual competence and motivation?

We will explore these issues through a series of five educational interventions in K-12 classrooms:

1. Having students play WebQuest games that have been authored by us or by the teachers.
2. Having students author their own WebQuest games for their peers to play.
3. Enhancing the use of factual questions and keyword answers in WebQuest with open-ended questions and essay answers, evaluated using automated LSA mechanisms.
4. Allowing students to share WebQuest games and game components across the Internet using Remote Explorium software that we have developed.
5. Supporting the creation and sharing of theme-centered sets of WebQuest games and related curricular resources using Teacher’s Curriculum Assistant software that we have prototyped.

These interventions and the evaluation of their effects will be described in Sections 4 and 5, following a discussion of the potential of LSA.

3. Cognitive Research

The proposed work is based on constructivist and collaborative theories of learning, broadly defined. The psychological background most relevant to the proposal is the construction-integration theory of text comprehension
(Kintsch, 1994) and the latent semantic analysis theory of knowledge acquisition and knowledge representation (Landauer & Dumais, in press). This theoretical framework is complementary to the cognitive theories guiding our design of computer support for learning: breakdown / repair (Fischer, 1994) and situated interpretation (Stahl, 1993). For the most part this background is relatively widely known; since the proposed project centers on the application of a technique that is less well known, we will focus on explaining latent semantic analysis in this section.

What is LSA?

Latent semantic analysis (LSA) is a mathematical / statistical technique for extracting and representing the similarity of meaning of words and passages by analysis of large bodies of text. LSA uses singular value decomposition, a form of factor analysis, to condense a very large matrix of word-by-context data into a much smaller, but still large/typically 100-350 dimensional/representation. (Berry, Dumais & O’Brien, 1995; Deerwester, Dumais, Furnas, Landauer & Harshman, 1990). The right number of dimensions has been discovered to be crucial; with the best values, which can be easily optimized for a domain, LSA yields up to four times as accurate simulation of human meaning judgments as ordinary co-occurrence measures.

The promise of LSA

Several sources of evidence show that LSA validly reflects human knowledge of word meaning and human interpretations of terms in text passages:

- After training on about 2,000 pages of English text, LSA scored as well as average test-takers on the synonym portion of TOEFL (the ETS Test of English as a Foreign Language).
- After training on an introductory psychology textbook, LSA equaled students' scores on a multiple-choice hour exam.
- LSA significantly improves automatic information retrieval in general by allowing user requests to find relevant text on a desired topic even when the text contains none of the words used in the query.
- The semantic similarity of successive sentences as measured by LSA mirrored manipulated variations in coherence in expository texts and accurately predicted their comprehensibility (Foltz, Kintsch and Landauer, 1994 ).
Proposals for Research

- Simple averages (centroids) of the words contained in these sentences significantly predicted the semantic priming by sentences of words judged to be related to the sentences’ overall meaning (Landauer and Dumais, in press).

- Pilot studies have found promising results of using LSA (a) to predict which of a set of brief texts an individual student will learn most from depending on the student’s prior knowledge as expressed in a short essay (research in progress), and (b) to evaluate the content of essays based on their LSA resemblance to text studied by the student or to pre-scored essays written by other students (Foltz, 1996, and research in progress).

Potential uses of LSA for learning and teaching

In what follows, we mention a number of examples of potential educational applications that appear worth pursuing eventually. Since we are proposing to develop a tool, it is important to form some idea about the possible range of uses for this tool. Of course, we can investigate only certain of these uses in the present project, as we shall describe in Section 4.

We believe that LSA can eventually provide the basis for a spectrum of effective new tools for facilitating and enhancing exploratory, project-based and collaborative learning, and we mostly describe such potential applications. However, we believe that most of the methods could also be applied in conjunction with other educational styles and methods, including computer-based tutoring, independent study, and traditional classroom instruction. In all cases, the goal of the new tools is not to supplant other methods, but to augment and amplify their benefits to learners and to help educators produce more and better learning with the same educator effort.

1. Finding optimal text for learning. Since actually finding relevant sources in a large information base such as a library or WWW is very difficult, the teacher traditionally provides a closed set of resources containing the necessary information. Furthermore, the problem is not merely to find resources relevant to a topic, but to find ones comprehensible to the learner with a particular background knowledge. LSA may be able to enhance a teacher’s ability to automatically match educational resources to individual students.

LSA is not only capable of selecting topic relevant materials, but it is able to match individual levels of prior knowledge and terminological sophistication as well. A research project directed by Landauer & Kintsch and funded by DARPA has shown that LSA can be used to choose, from a
set of texts on a particular topic, the one text from which an individual student will learn the most. The underlying principle is a notion adapted from Vygotsky (1968), “zones of proximal learning” (Kintsch, 1994). A student learns most from text that, on the basis of prior knowledge, is understood with moderate effort and contains just the right amount of new information. Students are first asked to write short essays on a topic, then the LSA centroid of their essays are compared with those of texts on the same topic but at varying levels of sophistication. Wolfe, Schreiner, Rehder, Landauer & Kintsch (in preparation), using texts about heart function, have shown that LSA-based choice of a text for an individual can result in about 50% more learning than random choice of text.

2. Coherence and comprehensibility measurement. LSA can be used to automatically measure text coherence and comprehensibility, important aspects of thinking and its written expression. Automatic evaluations could be incorporated as a component for a computer tutor in most subjects, or directly serve as an aid for independent learners/much like current spelling, grammar or style checkers.

3. Connecting students with each other and with relevant experts. LSA could also be used to match more effectively a particular student to other people with similar interests for conversation, collaboration or consultation. Students could either post messages on the Internet or leave statements of interest with characters in a WebQuest game. A computer-based agent would collect them, make LSA comparisons to match areas of interest or levels of knowledge, then pass on recommendations of people to get together with or automatically initiate interactions.

**Potential uses of LSA for educational assessment**

LSA can provide automatic ways to objectively evaluate written products and to generate content-customized, objective test items. It appears that it may be feasible to automatically measure, at least approximately, the following:

- The quality and relevance of individual written contributions to group activities.
- How much a student has learned from the materials that only she or he has read.
- The overall correlation of an individual’s contributions to the continuous process and final (textual) product of some kinds of collaborative group activities.
The point of such methods would not be to supplant the professional assessment skills and judgment of teachers. Rather, in face of the virtual impossibility of a teacher devising and grading equivalent tests for each student where each has studied a different, unanticipated subject matter, the intent would be to supplement and contribute to teacher judgment of overall achievement. To this end, LSA would be used by the teacher to produce and score a battery of brief assessment instruments individually targeted (a) to the idiosyncratic knowledge being acquired by each group of project participants; and (b) to the different knowledge sources encountered and activities engaged in by each individual.

1. Automatic writing assessment. Evidence that LSA can assess the quantity and quality of learned knowledge contained in a student’s writing has come from several kinds of studies. The most direct have been explorations of the use of LSA to automatically assign grades to essay exam questions. Predictions of instructor assigned grades were quite good, \( r = 0.67 \), and about the same as the correlation of 0.68 between two human graders.

Concretely, the application of LSA to assessing student knowledge and expression in exploratory and project-based learning might proceed as follows. At the beginning of a project, someone/the teacher, a publisher, a curriculum specialist, an independent student/would collect a large and broad training corpus of text relevant to the overall topic, by assembling electronic text either from textbooks and articles or by an Internet search, followed by some culling and editing, and submit it to LSA. As students found texts they would also be included. To evaluate a student’s knowledge and project contributions, an LSA-based program would be invoked by the teacher or student. It could be asked to perform one or more possible actions. For example, it might produce an estimate of the relevance to the overall topic of each text by a particular student (as always, with unusual pieces flagged for teacher attention). It might be asked to determine the similarity of a student’s computer-composed summary of research findings (or other communications and contributions to the group effort) to individual or group source material or to the group’s final report. It might be asked to score an answer to an essay question posed by the teacher, who might have devised the question after searching on a special subtopic among one or all students’ discovered sources. Note also that LSA could be used to detect instances of too much overlap with particular source materials, suggesting over reliance on a select and paste strategy in writing.

2. Choosing or constructing a summarizing sentence or paragraph. LSA may be able to order, and possibly quantify predictively, the quality of every
sentence in a text, and thus score a test item consisting of text on a desired topic from which test-takers are asked to construct a summary sentence.

3. Choosing or producing related concepts. LSA could be used to find related sets of words or phrases from a collection of texts on a topic and to estimate their similarity for concept matching or relating tests.

4. Portfolio assessment. It is conceivable that LSA could be applied usefully as a partial or component scoring technique for text-based portfolio evaluation. One idea would be to use LSA to measure the coherence of student generated text by measuring the semantic relatedness of successive sentences, as in the experiments mentioned earlier. Another idea would be to use LSA to measure the degree to which the text produced by students reflects the range of content available in the textual resources with which they have been provided or have selected on their own. Coherence in writing, together with topic relevance in comparison with source texts, not only reflects text-based understanding, but can also be taken as an indication of ability to successfully relate ideas, to reason and to transform knowledge.

We have mentioned here a broad range of conceivable applications of LSA to technology-enhanced education only to suggest the potential richness of the approach. In the next sections, we describe our specific goals for the present project.

4. Educational Intervention

We currently plan the following stages of intervention using our software with LSA in the classroom. Because of our commitment to user-centered design and student-centered activities, we will be responsive to what we observe in the classroom and to the suggestions and interests of students and teachers. Thus, the following plan will serve as a guide to help us focus on our research interests rather than a rigid recipe for our work during the duration of the project:

Stage 1. Students play WebQuest games. We have been exploring this stage in several K-12 classrooms during the past school year. We will continue to work with teachers and students to design new types of games and to use them differently in a variety of classrooms. Games we design can then serve as prototypical models to inspire students to construct their own games. Building different kinds of games also gives us insight into the usability of our software and ideas for new functionality.
Stage 2. Students author their own games for their peers to play. We have just begun to explore this approach and have already found that it makes a great deal of difference. Students not only construct their own knowledge of a topic in order to teach it to peers, they become engaged in a design process to structure the knowledge effectively. We believe that design skills provide important learning capabilities for the information-intensive future. Much of our computer science research has centered on developing computational media to support design, and our educational software takes design as a metaphor for constructivist learning. Thus, we have developed a series of software environments that support learning related to a task at hand (Fischer, Nakakoji, Ostwald, Stahl, Sumner, 1993). When students design games for other students, they engage in authentic, self-motivated tasks, reflect on their own or their peers’ learning processes, participate in important social interactions, and interpret domain concepts from different perspectives (Stahl, 1993). Classrooms in which students play each other’s games become involved in joint construction of knowledge.

Stage 3. Use of open-ended questions. This is where LSA is needed to create “the articulate classroom”. At this stage, game authors define answers to scroll questions by writing brief essays. Players then answer the questions with their own brief written responses. LSA mechanisms compare the two texts and judge whether they are sufficiently similar in content. This allows students to express their understandings in their own words. LSA is particularly effective in matching up different ways of saying the same thing using different vocabulary. Questions that required rote recitation of facts like the names of Jupiter’s moons can now be replaced with thought-provoking questions like: What would be the effect of Jupiter's gravity on a space ship that wanted to land on Jupiter? Writing paragraphs on such questions promotes high-level learning processes and develops scholarly communication skills.

Stage 4. Students share games across the WWW. Incorporation of supplementary software we are developing (the Remote Explorium and the Teacher’s Curriculum Assistant, described on the following pages) opens up the knowledge-building community to the world. The articulate classroom becomes a global classroom. A student who has developed a game on an esoteric topic can find other students interested in the same topic by distributing the game on the Internet. In this way, WebQuest games will provide yet another communication medium for students on the WWW. The distribution of games also creates a wealth of educational resources for teachers and students to choose from for their group and individual activities. This stage stresses the potential of the Internet to be an active two-way communication medium, rather than just a static repository of
information. Students learn to become actors in the scientific community, not merely consumers of external knowledge.

Stage 5. Theme-centered games incorporating written reports. The original WebQuest theme involves knights from the Middle Ages, deriving from the popular dungeons and dragons games. But WebQuest is built on a very general simulation construction substrate, so the visual appearance and the definitions of agents can be readily changed. To build a game on a new theme would be a major undertaking for an individual student. Although some students might want to do this for a theme they have already begun to explore, it makes more sense for a classroom to work together on this. The process might be as follows:

- The class selects a theme like the solar system. They begin researching the topic on the WWW to collect interesting WWW sites and stimulating questions.

- Students divide up the tasks of constructing background icons, character depictions, agent interactions. For instance, a WebQuest related to the solar system might include icons of the planets, spaceships, astronauts, cosmic ray dangers, space walk challenges, etc. Ambitious students could even build simulations into their games, including, for instance, graphic demonstrations of the effects of different gravitational forces.

- Individual students or small groups design games incorporating the components of the themes.

- Students play each other’s games and increase their knowledge of the subject matter.

- Students critique each other’s game designs and revise their own games.

- The class gets together to reflect on the experience, to discuss what they learned about the topic and to compile reports on the theme.

- The class shares what they have learned by distributing some of their games on the Remote Explorium. They might construct their own WWW site on the theme, including statements of their ideas and pointers to other sites they discovered.

Stage 6. Versions of questions and information sources for people with different background knowledge levels. People construct new knowledge by going beyond their previous understanding and then integrating the new insights into their background knowledge (Kintsch, 1994; Fischer, 1994; Stahl, 1993). Therefore, educational information is most effective for an individual when it falls within the person’s zone of proximal learning. LSA
allows us to personalize information sources to students by finding texts that most closely match (or slightly exceed) the student’s own writings. Rich digital libraries can provide selections of alternative presentations on any given topic. For instance, the Remote Explorium could eventually contain many versions of solar system games. These versions could be rated using LSA and indexed in the database of the Teacher’s Curriculum Assistant so that teachers and students could select the most appropriate versions. In addition to selecting entire games, people could find game components such as background features, character agents and question narratives. The Visual AgenTalk programming language used to define character behaviors is user-extensible and students could exchange little subroutines in this language that accomplish interesting interactions. LSA databases could also be exchanged across the Internet. That is, one classroom could collect and author texts on a particular subject, then submit it to LSA to create a database of interrelated terms. These databases can be used by other classrooms to evaluate student essays on the given subject, resulting in ratings of the students’ background knowledge and readiness to learn from resources at different levels.

By the sixth stage/which we plan to explore in the second half of the proposed project/teachers and students have a wealth of resources organized into coherent curricula on interesting themes. The resources are available in alternative versions for people’s levels of knowledge acquisition, or as constituent components that students can combine in their own constructions. In addition, there are software environments to support the collaborative construction of knowledge using these resources, including mechanisms for evaluating text and matching it to individual learners automatically. These tools will help teachers in their new roles, freed from some of the tedious evaluation of rote tests. They will have to oversee the progress of students and make sure that LSA ratings stay on track, using this information to judge what kinds of high-level guidance and support to provide. Our research will look at how to make most effective use of both teachers and software in the classroom.

The software we are planning to test and refine in classroom use in order to support “the articulate, global classroom” consists of the following three component systems currently being developed in our labs:

- **WebQuest**, a software environment for the design of educational games.
- **Remote Explorium**, a WWW site with WebQuest games and other educational simulations that can be down-loaded by users around the world.
Teacher’s Curriculum Assistant, a software environment for teachers to locate, evaluate, adapt and share educational resources over the Internet.

**WebQuest** (Perrone, Clark, Repenning, 1996) is an adventure game development environment we developed to research educational software like Carmen Sandiego. It allows a game author to lay out a graphical scene with fields, paths, lakes, islands, etc. The scene can then be populated with active agents, such as heroes, princesses, dragons, locked doors and buried treasures. (See Figure 2.) Scrolls are defined and associated with game obstacles. The scrolls pose questions that a player must correctly answer to get past a dragon or enter a door. The scrolls may suggest WWW sites to explore to find hints and answers to scroll questions. When a player clicks...
on a suggested site, the software opens a WWW browser displaying that site. Players can browse the WWW or perform WWW searches using standard search engines.

In a typical WebQuest game, a player might adopt a medieval knight character and be confronted by an anachronistic question like: What are the names of the four largest moons of Jupiter? The student would read a WWW page about the solar system, answer the question and then pursue the dragon. The question might also be one that requires more understanding and research, like: What was the right ascension of Mercury during the signing of the Declaration of Independence? (See Figure 1.)

The WebQuest software has the capability to let students construct original educational games for their fellow students. This creative process allows students to explore information on the WWW in self-directed ways and to embed ideas and facts they discover into game boards that they design. Students learn new information while situated within a context of having to incorporate the new information into the conceptual framework of an educational game they are constructing for their peers. Within a particular classroom, students exchange and play games, learning subject matter that has been organized by their peers and providing feedback to the game creators. Both game players and authors develop research skills using the WWW; they also both reflect on the organization of knowledge and the strategic design of the game artifact.

The authoring capability of WebQuest takes advantage of Agentsheets (Repenning, 1994), the programming substrate that WebQuest is built upon. Agentsheets is a substrate we developed for building educational simulation applications. It allows authors to design the appearance and behavior of their own active agents, as well as creating their own backgrounds with which the agents interact. Agentsheets is programmed by game authors entirely through visual manipulations and requires no traditional programming knowledge. It includes an end-user programming language, Visual AgenTalk (Repenning 1995), which allows students to define the behaviors of their agents. We have begun testing the Agentsheets and Visual AgenTalk authoring capabilities in the classroom with very positive responses. Students are enthusiastic about tools that empower them to construct their own software environments. The proposed project will allow us to pursue this research and to enhance it with the capability to analyze the process of knowledge building as evidenced by students' question and answer formulation.

The Remote Explorium (Ambach, Perrone, Repenning, 1995; Stahl, Sumner, Repenning, 1995) allows game authors to share their artifacts with
students elsewhere across the WWW. Teachers and students can download and adapt entire games or their constituent components from the Explorium. This software was originally developed by us to facilitate the distribution of educational applications written in Agentsheets. As part of the proposed project, we will extend the Remote Explorium to allow students in different classrooms and different schools to share their WebQuest games over the Internet. Currently, researchers at the University of Colorado can put Agentsheets applications on a WWW page for students elsewhere to download easily. To implement our vision of WebQuest as a collaborative learning project, we will have to extend the Remote Explorium to allow students to post their games to the Web, so that the sharing is bi-directional. We also envision people trading components of games, such as graphical depictions of characters, programmed agent behaviors, or collections of narrative questions and answers related to given themes. This allows students to design their own games while taking advantage of components created by other students. However, experience with Remote Explorium to date demonstrates that teachers and students need additional support to take advantage of the distributed resources. We have designed another program to provide just such support.

The Teacher’s Curriculum Assistant (Stahl, Sumner, Owen, 1995) retrieves summary information about games in the Explorium and elsewhere on the Internet. It uses this information to help teachers or students locate the game and curriculum examples on the WWW that best match their pedagogical needs. In addition, it provides curriculum ideas and resources to guide the classroom use of the games.

The problems that teachers have using the Remote Explorium are typical of the plight of people trying to obtain educational resources from the Internet generally:

- There are no effective methods for locating relevant curriculum sites, such as WWW pages containing WebQuest games on specific themes.
- It is difficult to search for items of interest; search engines are too generic and indexes to education sites are too idiosyncratic and anecdotal.
- There is no choice of versions for different ability levels, or if there is it is not systematically organized.

Figure 3. The Teacher’s Curriculum Assistant interface for locating, searching and selecting resources: the Profiler, Explorer and Versions. These tools help a teacher
find educational resources on the Web that are relevant to classroom goals and plans.

- There are no software support tools for adapting resources to one’s particular needs.
- There are no aids for organizing selected curriculum resources into coherent curriculum plans.
- There are no simple mechanisms for teachers and students to share their experiences by posting comments or new games back to the Internet.

Figure 4. The teacher-client interface for adapting, organizing, and sharing resources and curriculum: the Planner, Editor and Networker. These tools help a teacher incorporate educational resources into personalized lesson plans and share successes with these resources on the Internet.

We have prototyped a curriculum development design environment to respond to these problems. The Teacher’s Curriculum Assistant maintains a database of information about on-line educational resources. It uses information in the database through six user-interface components: Profiler, Explorer, Versions, Editor, Planner and Networker. The Profiler defines the user’s needs in order to query the database for relevant resources. The Explorer allows a user to browse among related resources and curriculum ideas. The Versions component explains the differences between different versions of the same resource so that the most appropriate one can be chosen. (See Figure 3.) The Editor is used for adapting resources (e.g., editing a text document. The Planner helps a teacher to arrange resources into a lesson plan and to make adjustments to the plan. Finally, the Networker simplifies Internet access, facilitating the posting of comments and new games as well as handling the downloading of selected resources and the updating of the database. (See Figure 4.)

The Teacher’s Curriculum Assistant was designed based on our philosophy of adapting curriculum and resources to the particular pedagogical needs, learning styles and personal interests of the students and teachers in a classroom. The proposed project will allow us to explore the use of LSA in matching textual materials to the background knowledge of individual students, taking full advantage of the built-in support for multiple versions of resources.
5. Experimental Design

We are developing a suite of educational tools in continuous interaction with classroom experience with the tools. We are not designing a finished product, to give to a teacher and evaluate how it works. Rather, we start with prototypes that have some of the features we think we eventually want to have, obtain feedback about their performance, and gradually modify and elaborate our designs. As we have pointed out above, it is not so much the software tools that we are concerned with, but how they are to be employed effectively in the classroom.

The current version of WebQuest has been used in Boulder middle schools. We intend to expand this use locally, including into high schools. We have close working relationships with a number of teachers in different schools who have used our software in their classrooms and who are eager to try WebQuest. In addition, we have contacted several non-local groups and made plans for possible future cooperation. We would let those groups use our tool in the way they prefer, but obtain data on their project from them as well as collect our own data by sending a project member to visit and observe out-of-town sites at regular intervals.

The kind of data we plan to collect are both observational and experimental. Observational data will come from teachers, students, project members who observe classroom use, along with records automatically collected by the computer systems themselves. A project diary will be used to help us organize and preserve these observations and permit their use at later points in time. While much of the observational data will necessarily be informal and opportunistic, we also plan to develop organized observation protocols to ensure comprehensibility and facilitate comparison. The construction of such a protocol would be one of the research goals for the first project year.

We do not plan any large scale classroom evaluation experiments, which would be premature as well as exceeding the resources of the project. Instead, mini-experiments directed at specific questions that arise in the course of this project will be used. At this point, we can sketch only a few obvious first experiments, but these should make it clear how future experimental and evaluation research in this project could proceed:

- Is WebQuest effective as a tool for learning how to search the WWW? Groups of students with varying amounts of experience using the WWW with WebQuest are compared with equivalent students using the WWW with traditional "how to" instructions. Their success at specified search tasks as well as browsing behavior and browsing strategies will
be evaluated. Follow-up questionnaires can be used to assess long-term effects.

- How well does LSA evaluate written student responses? We are hopeful that LSA can make fine enough distinctions to identify plagiarism/strong temptation when students can easily cut and paste from WWW pages into their essays. We will have to determine empirically whether scores of very high similarity between an essay and a resource text indicate the likelihood of literal copying. Throughout the project, specified samples of student responses will be scored both by LSA and human graders to evaluate the effectiveness of LSA to evaluate written student responses.

- How effective is LSA in helping students to formulate questions? The precise experiment cannot be outlined at this point, because it depends on just what we shall come up with in this regard and how the WebQuest components will evolve. But eventually formal experimental comparisons can be made not only between our support system versus no support, but also between a teacher-led group discussion and the LSA support system. Such studies would be important not so much because they might tell us that LSA is 50% or 70% as effective as a good teacher, but because it might pinpoint differences in the way a teacher helps and the ways our system can be used. Data like this could be more reliable than informal and fortuitous observations and would help direct the evolution of our system.

- During the use of LSA tools, we can easily and automatically make them available or not to a particular student working on a particular topic or question. By randomizing these assignments and using LSA-based custom evaluation tools we will be able to do almost continuous objective measurement of the effect of the tools on problem solving and knowledge acquisition. Statistical analysis by classical randomized within- and between-subject differences (simultaneously in this design) will be straightforward.

- How can software environments best be used in the classroom? Constructivist approaches like game creation typically require longer time commitments and more individualized work than traditional school schedules can accommodate easily. Solutions to this problem will be investigated by working with teachers and trying different ways to integrate the use of the software into classroom processes. We will try small group projects, independent student efforts, after-school arrangements, etc. in order to allow motivated students to develop exceptional but time-consuming games. We will explore different ways
of sharing work among groups, between classes, having classes build on previous year’s accomplishments. WebQuest will be introduced into a range of schools, from more traditional to more experimental to see how different solutions can be found in different organizational contexts.

- How can coherence of knowledge be promoted? Without guidance, students authoring WebQuest games will tend to build an unstructured sequence of questions and answers. One of the goals of evaluation will be to examine this possibility and to determine what kinds of constraints can be built into the system so that students construct coherent bodies of knowledge. Although resolving arbitrary relationships between a game situation and questions posed may be an impetus to students' creativity, deeper learning will result if questions build on each other, and motivation may be better sustained if the questions are related in a meaningful way to events in the game. For example, the discovery of an underlying relationship in pieces of topic knowledge encountered might become a goal for progressing through the game. LSA may also be useful here in helping students construct an interrelated network of concepts and ideas from the information collected from multiple sources.

- How can software environments best support learning? In addition to providing challenges and sources of information, software can provide guidance. For instance, the LSA mechanisms can be used to guide students to the most appropriate versions of materials. When software mechanisms determine that a student response is inadequate, they can suggest further sources of information to be consulted. We have used computational critics in many of our other software environments to alert users to relevant information (Fischer, Nakakoji, Ostwald, Stahl, Sumner, 1993), and will try to combine critics with LSA tools in this project. We will observe how effective these techniques are within classroom practice.

- How can the Internet be used as a medium for the collaborative construction of knowledge? In the later years of the project we will investigate the effectiveness of tools like the Remote Explorium and the Teacher’s Curriculum Assistant in turning the WWW into a bidirectional medium in which students contribute knowledge as well as consume it. It is premature to determine how specific functionality of this software will be evaluated.

- Where is the best boundary between what must be known and what can be found when needed? In general, we want to further our understanding of an important conceptual problem that must be dealt with if
technology such as that proposed here is to be used effectively in education. The problem is the relation between external memory and internal memory. Without some internalized knowledge, external information sources cannot be used effectively. Over-dependence on external memory may discourage the construction of internal knowledge. Clearly, one cannot teach all the knowledge a person might at some point need. Is there something one must know in order to be able to understand what one looks up, and if so, what is that essential knowledge or skill that we need to teach? To what extent does this involve general knowledge? To what extent is it tied to specific domains? Our Web-based information retrieval capabilities will be excellent in a few years/we need to make sure that our understanding of the conceptual issues concerning the knowing / finding tradeoff keeps up with our technological capabilities.

The primary responsibility for evaluation methods will rest with Drs. Tom Landauer and Walter Kintsch, two experimental psychologists with a great deal of experience in research like this. They will also be the LSA experts of the project. The development of the WebQuest system and its integration with Remote Exploratorium and Teacher’s Curriculum Assistant will be under the direction of Dr. Gerhard Fischer, a computer scientist experienced in software research, in cooperation with Dr. Gerry Stahl and Corrina Perrone. The task of integrating WebQuest into classroom activities will be directed by Dr. Eileen Kintsch in cooperation with David Clark, a Boulder teacher now using WebQuest who is an authority on student use of the Internet (Clark, 1995).

References


Part IV: Other Proposals at the University of Colorado
Perspectives on Collaboration: A Micro-ethnographic Study of Computational Perspectives in Computer Support for Collaborative Knowledge Building at a Virtual Biology Laboratory

PROJECT SUMMARY

Collaborative Knowledge Building

This project undertakes a study of small groups constructing scientific understanding as a process of collaborative knowledge building. It designs, develops, deploys, and studies computer support for such collaboration, especially under conditions in which participants are not co-located and cannot engage in face-to-face interaction. Collaborative knowledge building is conceived as a set of related activities through which a group develops a gradually deepening understanding of some area of inquiry. It contrasts both with cognitivist views of learning focused on the individual and with support for exchange of personal opinions or short-term decision-making.

Computational Perspectives

A key structure of collaborative knowledge building is perspectives. Collaboration proceeds largely through the making and taking of personal and group perspectives. This project investigates the support of such a knowledge building structure through the use of “computational Perspectives,” which represent or support the evolving network of personal and group perspectives. The central research question is whether participants in computer-mediated collaboration can effectively and intuitively make use of a computational Perspectives mechanism.

Cognitive Artifacts

The ability to engage in scientific knowledge building is dependent upon the ability to understandingly use a variety of scientific artifacts. Artifacts such
as simulations, analysis tools, and data sheets significantly extend the power of native human cognitive abilities; they also serve as persistent communication media to express and preserve insights for others. This project conceptualizes computer support systems as sets of cognitive artifacts. It conducts micro-analyses of how people – individually and as communities – develop a practical understanding of these artifacts.

**Micro-ethnographic Analysis of Interaction**

Micro-ethnography is a rigorous social science that incorporates recent methods and findings from the analysis of verbal and visual human interaction. It uses digitized video to study interpersonal behavior at a detailed level. Micro-ethnography will be used in this project (a) to analyze the structure of negotiation in small group meetings in order to design software support of negotiation of ideas in personal and group Perspectives; (b) to study how people learn to use computer-based and internet-based cognitive artifacts that are part of a virtual biology laboratory; (c) to study specific distance-collaboration software as effective media for supporting perspectives in knowledge building.

**Study of a Virtual Biology Laboratory**

The project will ultimately study collaborative knowledge building at a virtual biology laboratory used in geographically distributed high schools for advanced placement biology courses. A website containing the virtual biology lab is currently being developed at the University of Colorado; this project will assess its effectiveness in use by college freshmen and contribute to its iterative design (in Year I). The project will also develop a collaborative knowledge-building environment (in Year I) and integrate this with the lab (in Year II). Then (in Year III), micro-ethnographic methodology will be used to assess the use of this software – especially the use of computational Perspectives – by distributed high school students, and contribute to its iterative design.

**Building on Previous Research**

The project PIs and Advisory Board members have conducted research projects (including a three-year CSS grant) and specific pilot projects that form a foundation for the proposed research. The PIs have previously developed knowledge-building environments with computational Perspectives for designers and students, have studied the theory of cognitive artifacts, and have engaged in micro-ethnographic analysis of students
learning to use a scientific simulation. This and related work by others motivate new features and research issues. A previous Perspectives system will be extended with additional functionality and re-structured for integration with the virtual biology lab, for release as Open Source, for use in distance collaboration, and for micro-ethnographic analysis.

PROJECT DESCRIPTION

1. Collaborative Knowledge Building

This project undertakes a study of small groups constructing scientific understanding as a process of collaborative knowledge building. It designs, develops, deploys, and studies computer support for such collaboration, especially under conditions in which participants are not co-located and cannot engage in face-to-face interaction. Collaborative knowledge building is conceived as a set of related activities through which a group develops a gradually deepening understanding of some area of inquiry. It contrasts both with cognitivist views of learning focused on the individual and with support for exchange of personal opinions or short-term decision-making.

Learning as Collaborative Knowledge Building

The term “learning” can refer to a wide range of phenomena, from the accumulation of facts to the development of deep understanding; from the experience of something new to the mastery of specific complex skills. Attempts to assess learning range correspondingly widely, from assumptions that people are always learning to frustration that people cannot transfer what they have learned to new circumstances (Russell, 1999). For the assessment of learning in the sciences, we prefer the term “knowledge building” (Bereiter, 2000). This refers to the progressive creation or construction of knowledge about a specific topic of inquiry (Dewey & Bentley, 1949/1991). For instance, a group of students may gradually propose tentative answers to a scientific question, accumulate relevant data, debate alternative arguments, and converge toward a deeper understanding of the phenomenon (Donald, 1991). This process of knowledge building takes place within a community of inquiry, in which experiments can be replicated, assumptions questioned, and insights discussed (Scardamalia & Bereiter, 1996). Knowledge building is an interpersonal process that can be observed and documented (e.g., using the micro-ethnographic methods described below), unlike learning, that is generally taken to be a
psychological phenomenon that can at best be inferred indirectly through tests or other outcomes (Lave, 1991).

Knowledge building is inherently collaborative. By “collaborate” we simply mean “to work together” (from the Latin com laborare). To build knowledge is to formulate theories and similar bodies of knowledge that are (or at least could, in principle, be) shared by a group of people and that meet criteria for contributions to knowledge that are accepted within a scientific community (Latour & Woolgar, 1979). A high school biology class, for instance, must develop knowledge of biological phenomena and theories that meet criteria that gradually approach the standards of the field of biology. Although knowledge may necessarily involve the minds of individuals, it is generally the result of interactions with other people, with cultural artifacts, and with shared language (Vygotsky, 1930/1978). It is therefore often useful to conceptualize and analyze learning as distributed across “units of analysis” or activity structures that include small groups, their tools, and their language (Engeström, 1999; Hollan et al., 2000; Hutchins, 1996).

This project is guided by a theory of collaborative Knowledge Building Environments (KBEs) that we are developing (Stahl, 2000c). The theory proposes the following principles:

• Collaborative knowledge building is a particular view of group learning that focuses on a range of activities that take place within communities, as opposed to focusing on learning as the transmission of bits of information to individual learners.

• Collaborative knowledge building takes place largely through the interaction among people with different understandings from multiple personal and group perspectives.

• Such knowledge building within groups can be helped by appropriately designed computer technology that supports various knowledge-building activities and supports interaction among alternative perspectives.

In the following sections, we discuss our model of knowledge building, the potential of computer support, and the role of perspectives.

**A Model of Collaborative Knowledge Building**

One approach to better understanding how to design computer support for collaborative knowledge building in social settings is to conceptualize the various constituent activities involved in individual and social knowledge building. Figure 1 from (Stahl, 2000c) provides a starting point for this,

The idea of this diagram is that knowledge building can proceed through many different activities. The sequential structure of the model is only illustrative of a typical activity series. We understand that these activities overlap in practice. The possible relationships among the individual activities – and particularly the interactions between the personal and social – can be complex and varied. The purpose of the diagram is to suggest a number of distinct activities that could be supported by software with multiple functionality; the sequential flow is not intended to imply a necessary order to the activities.

Figure 1. A model of personal understanding and social knowledge building.

A set of seminal books and articles in computer-supported collaborative learning (CSCL) has formulated a view of learning as a social process of collaborative knowledge building within communities of practice (Brown & Campione, 1994; Brown & Duguid, 1991; Koschmann, 1996; Lave, 1991; Lave & Wenger, 1991; Pea, 1993; Scardamalia & Bereiter, 1996; Wenger, 1998). However, these texts do not make the set of cognitive and social activities that underlie such a view explicit in the manner attempted in our theory. Starting in the lower left corner, Figure 1 shows a cycle of personal understanding. The rest of the diagram depicts

how personal beliefs can be articulated in language and become part of social interaction. Note that the results of social knowledge building
eventually feed back into personal understanding, providing the evolving toolkit of culturally-based individual cognitive capabilities. The collaborative knowledge-building process begins with (a) the articulation in language and (b) the confrontation of these statements with (c) alternatives from other perspectives. The interplay of perspectives proceeds through various interactional mechanisms, potentially culminating in the reduction of shared knowledge to a text or other persistent artifact (d-k). Computer support for collaboration could support many of the activities represented in the model, including the roles of perspectives.

The Potential of Computer Support

Based on our own experiences with software in classrooms, we have found that computer-supported collaborative learning has a vast – and largely untapped – potential (Kintsch et al., 2000; Stahl, 2000d; Stahl & Sanusi, 2001; Steinhart, 2000). Access to global sources of information is just one facet. In addition, computer simulations can transform conceptual representations into interactive worlds for inquiry. They can transcend real-world barriers of time, expense, geography, scale, expertise, etc. to allow students to engage with and experience phenomena that have until now been unapproachable – such as Nobel prize-winning biology experiments. Hypertext systems of information can personalize presentations to meet individual learning needs, providing links to both remedial and supplemental information. Communication media can promote collaboration in ways never before possible, as well as among people who could not hitherto interact, allowing students to work with other students with similar interests far away. Structured curricular databases and shared knowledge-building environments can support student learning processes by providing access to ideas of scientists and fellow students in persistent forms that can be thought about and inter-related.

However, we have seen that people always use computer artifacts in ways not envisioned by the designers. So, careful study of the artifacts in naturalistic settings is critical to the development of effective collaboration technology. This project will develop and integrate a VIRTUALBIOLOGYLAB Web-based system that embodies these potentials, and will study the degree to which these potentials are achieved in practice.

We have also seen from extensive studies of experimental knowledge-building environments by other researchers – like CSILE/KNOWLEDGEFORUM (Scardamalia & Bereiter, 1996), KIE/WISE (Cuthbert, 1999), and COVIS (Pea, 1993) – that collaboration support for
Learning can be powerful. Environments like these transcend rote learning of isolated facts by engaging groups of students in discussions and explorations of challenging and meaningful scientific issues. They provide more than generic chat windows to encourage students to engage in scientific inquiry through discussion with other students, and scaffold the exploration of leading scientific themes. Like anyone developing scientific understanding, students should also have knowledge management tools to organize, categorize, revise, summarize, question, and propose. If they are to learn how to approach a topic using scientific modes of thought that are new to them, then their interactions with other students should be scaffolded and supported. We will develop WEBPERSPECTIVES software that supports collaborative knowledge-building activities, and we will integrate WEBPERSPECTIVES with VIRTUALBIOLOGYLAB, incorporating some of the scaffolding mechanisms explored in the systems mentioned earlier in this paragraph.

**Knowledge-Building Environments**

The form of computer support that we are interested in – a collaborative Knowledge-Building Environment (KBE) – represents a distinctive approach that overlaps related work in educational technology, computer-supported collaborative learning (CSCL), and computer-supported cooperative work (CSCW). Support for learning has traditionally been oriented toward the transmission of information to individual students. Even where it is based on a view of student construction of knowledge, as with Intelligent Tutoring Systems (ITS) for algebra or physics, the goal is measured by testing the incorporation of pre-defined content or methods into the individual’s understanding (Wenger, 1987). A more student-centered, constructivist approach is taken by Interactive Learning Environments (ILE), that might, for instance, allow students to create ecologies in SIMLIFE to learn biology, or programs in TURTLE LOGO to explore math concepts (Papert, 1980). In contrast, a KBE primarily supports the group process and leaves matters of content up to the participants (that may include a teacher who raises particular content issues and helps maintain focus, or a website with content and scaffolding). In this way, it applies CSCW approaches to CSCL.

A review of CSCW technology for groups (Kraemer & Pinsonneault, 1990) distinguishes group communication support systems (GCSSs) from decision support (GDSSs). GCSSs are generic communication media like email, chat, threaded discussion, and video-conferencing. KBEs need functionality that is more specifically designed to support knowledge-building activities.
GDSSs add tools for specific types of group interactions (e.g., voting) in providing computational tools for group decision making, but tend to support isolated, focused activities that collate the work products and opinions of individual members of a group. Whereas knowledge building is generally an open-ended evolution, GDSSs focus more on supporting short-term, well-defined decisions. In contrast, a KBE aims to support a broad spectrum of knowledge-building activities – such as activities (a) to (k) in our model – in a way that allows deep knowledge to evolve and emerge over time. Rather than just exchanging participants’ existing personal opinions, it supports the construction and interaction of alternative formulations of knowledge. It also supports the interplay of individuals and groups more comprehensively, through integrated mechanisms of divergent computational Perspectives and convergent negotiation processes that treat a group as more than just the sum of its individual members.

Assessments of CSCL and CSCW systems have defined a number of key issues for evaluating the problems and successes of such systems. For instance, in simple threaded discussion forums common problems include: short threads (a tendency for discussions to die quickly), low participation (lack of motivation to participate), few cross-references (little convergence of ideas), and superficial content (minimal depth of investigation) (dePaula, 1998; Guzdial & Turns, 2000; Hewitt & Teplov, 1999). On the other hand, GDSSs and GCSSs have been shown to decrease communication barriers within the group, while increasing task-oriented focus, depth of analysis, and decision quality (Connolly, 1997; Kraemer & Pinsonneault, 1990). Social informatics studies raise additional issues of software deployment and adoption as well as questions of usability and utility (Kling, 1999). These are some of the dimensions along which we will assess our software within naturalistic learning and working social contexts.

In summary, we want to design, develop, and assess KBEs that go beyond generic chat and discussion systems (that tend to encourage exchange of personal opinions or isolated facts, but not deep shared understanding and critical inquiry). Such systems should include specific tools and structures to promote on-going debate, knowledge management, and group decision-making; however, knowledge building should go beyond the management and dissemination of existing knowledge to support the emergence of qualitatively new, increasingly shared knowledge within a community (Engelbart, 1962; Engelbart, 1995). The following sections discuss tools to support this within the proposed project.
2. Computational Perspectives

A key structure of collaborative knowledge building is perspectives (Stahl, 1993a). Collaboration proceeds largely through the making and taking of personal and group perspectives (Boland & Tenkasi, 1995). This project investigates the support of such a knowledge-building structure through the use of “computational Perspectives,” that represent or support the evolving network of personal and group perspectives. The central research question is whether participants in computer-mediated collaboration can effectively and intuitively make use of a computational Perspectives mechanism.

Perspectives in Knowledge Building

Our theory claims that collaboration centrally involves interaction among multiple personal and group perspectives. According to the philosophy of interpretation (hermeneutics) human understanding is fundamentally perspectival. We construct knowledge from our situated perspective in the world: our historical position, cultural tools, and practical interests (Gadamer, 1960/1988; Habermas, 1981/1984; Heidegger, 1927/1996; Stahl, 1975). Computational support for knowledge building can represent our interpretive perspectives with computational Perspectives (Boland & Tenkasi, 1995; Nygaard & Sørgaard, 1987; Winograd & Flores, 1986). (In this proposal, Perspective—with-a•capital-P refers to the proposed computational mechanism that mirrors human interpretive perspectives—with-a•lower-case-p.) In this sense, Knowledge-Building Environments (KBEs) with computational Perspectives are designed to support the essential structure of collaboration. A key working hypothesis of the proposed project is that KBEs benefit from an approach that represents the perspectival nature of collaboration. A goal of the research is to facilitate the incorporation of a computational Perspectives mechanism in KBEs and study its use in settings of distance collaboration.

Computational Perspectives have been explored by the PI in a number of software prototypes, in his dissertation system, and in his theoretical publications (Fischer et al., 1993a; Fischer et al., 1993b; Stahl, 1993a; Stahl, 1993b; Stahl, 1998; Stahl & Herrmann, 1999; Stahl et al., 1995a; Stahl et al., 1995b). In a single-user system, computational Perspectives may correspond to different domains or professional viewpoints on a design problem, such as electrical, plumbing, structural, and heating concerns in architecture (Fischer et al., 1993b). In a KBE to support collaboration, computational Perspectives typically provide personal or group workspaces for the development of different sets of ideas. In this way, they can model the
Proposals for Research

relationships among the various personal and group interpretive perspectives at work in the construction of collaborative knowledge. This project will introduce computational Perspectives into distance collaboration for the first time.

The project will develop a KBE with support for personal and group perspectives. This WEBPERSPECTIVES software will extend functionality we have already developed and deployed in WEBGUIDE (Stahl, 2000d), adapting its architecture for distance collaboration systems. Most communication, conferencing, and collaboration media provide no support for organizing contributions according to who made them, or for building group perspectives from selections out of personal perspectives. Some systems provide two fixed levels: personal and group – either by limiting access to various pieces of information or by defining personal and group workspaces. WEBPERSPECTIVES will be the first system that enables multiple levels of groups and subgroups, and allows new subgroups to be added interactively. Moreover, WEBPERSPECTIVES automatically includes information from certain perspectives in other perspectives (according to a user-defined “content inheritance” lattice of perspectives), so that all information accepted by a group is incorporated in the perspectives of that group’s subgroups and members. This provides a computationally supported knowledge management and knowledge-building approach appropriate to the structure of interpersonal collaboration.

Software Support for Perspectives

The core of WEBPERSPECTIVES is a Perspectives server that queries a database of textual contributions to the ongoing discussion and provides user interface clients with the notes that are to be displayed in the Perspective requested by the user. The Perspectives associated with a particular knowledge-base form a non-cyclical lattice (each Perspective may have multiple parents or super-groups and multiple children or subgroups). The knowledge-base for a biology class would typically have a Perspective for questions and ideas shared by the whole class, several subgroup Perspectives for teams of students who work together, a personal Perspective for each student, and comparison Perspectives that bring together contributions from the Perspectives of all members of a group. A Perspective defines an electronic workspace in which a person or group can develop ideas and manage information that belongs together – for instance because it represents the beliefs and viewpoint of a particular person, group, domain, or intellectual position. Perspectives structure a shared information
space so that special coherent views can be built up and displayed. The mechanism of computational Perspectives is very general and flexible.

The design philosophy behind computational Perspectives as implemented in WEBPERSPECTIVES is that users have complete control over the content in their personal Perspectives. Thus, if my personal Perspective inherits conflicting ideas from different team Perspectives that I belong to, I can delete, edit, and rearrange those ideas at will – without affecting how those ideas appear in other Perspectives. Other users can view the contents of my personal Perspective (except for content that I have designated as private) and they can copy items, link to them, initiate public discussions of them, and propose them for incorporation in team Perspectives – but none of this affects how the content of my Perspective is displayed to me. This allows me to build my own Perspective on the topics that are under consideration by the group. I can see what knowledge others are building, incorporate that knowledge into my Perspective, or join in with others to share, discuss, and negotiate. The same design philosophy applies of course to team Perspectives: team members jointly (through negotiation processes) have complete control over the content of their team Perspective.

Inheritance is a central defining mechanism of computational Perspectives as used in this proposal. The ability to define arbitrarily complex networks of Perspectives with multiple layers of sub-groups between the group Perspective and the individual personal Perspectives, and to have the automatic inheritance of content through the network distinguishes this approach from all other systems of “views” and “perspectives.” Inheritance in this sense is not class inheritance, but “content inheritance.” A given Perspective can inherit content from multiple other Perspectives. This content is aggregated (logical union) in the given Perspective, where it can be over-ridden with edits, deletions, rearrangements, virtual-copying (linking), and additions. The inheritance mechanism is derived from efficient approaches explored in hypermedia, including “delta memory” and “transclusion” (Boborow & Goldstein, 1980; McCall et al., 1990; Mittal et al., 1986; Nelson, 1981; Nelson, 1995). For a discussion of related work, see (Stahl & Herrmann, 1999).

Because new Perspectives can be defined (either in advance or during system use) to inherit from any (non-cyclical) other Perspectives, it is generally useful to define “comparison Perspectives” that aggregate the ideas from team members, including those ideas that have not been agreed upon and migrated to the team Perspective. This is handy for keeping an eye on what one’s fellow team members are thinking. Typically, we have set up the inheritance network of Perspectives to have a diamond-shaped profile,
Proposals for Research

diverging out from the total group Perspective via teams to all the personal Perspectives, and then converging back via team comparisons to the group comparison Perspective. This models a collaborative knowledge-building process that combines divergent brainstorming and convergent negotiation.

3. Cognitive Artifacts

The ability to engage in scientific knowledge building is dependent upon the ability to understandingly use a variety of scientific artifacts. Artifacts such as simulations, analysis tools, and data sheets significantly extend the power of native human cognitive abilities; they also serve as persistent communication media to express and preserve insights for others. This project conceptualizes computer support systems as sets of cognitive artifacts. It conducts micro•analyses of how people – individually and as communities – develop a practical understanding of these artifacts.

Mediated Cognition

We start from three principles enunciated by Vygotsky (1930/1978; 1934/1986):

1. Mediation by artifacts. Modern human cognition is thoroughly mediated by physical and symbolic artifacts such as tools and words. We extend this to the use of computer-based artifacts like simulations, data analysis tools, and collaboration media.

2. Social cognition. Meanings and practices are first established interpersonally and may then be internalized in individual minds. We take advantage of this by analyzing the interpersonal interactions, that are largely observable to the trained analyst as well as to the participants.

3. Zone of proximal development. A student learns most productively when guided somewhat beyond his or her current developmental level by peers or a mentor. We use this principle to design experimental situations in which a small group of students is challenged to engage in a scaffolded scientific task.

The Role of Artifacts

It is possible to re-conceptualize learning (both individual and collaborative) through a focus on the artifacts that are involved. Artifacts – including software artifacts – embody intentionality, meaning, and experiences of their
creators and preserve these for future users (Donald, 1991; Hall, 1996). The problem is for users of artifacts to know how to reactivate this stored wisdom. This requires complex skills of interpretation (Gadamer, 1960/1988; Stahl, 1975; Stahl, 1993a; Winograd & Flores, 1986). Education can be viewed as largely the effort to socialize children and other newcomers into a practical understanding of the artifacts and practices that constitute a society’s or a community’s culture (Lave & Wenger, 1991). The written word and the symbols of mathematics, for instance, are cognitive artifacts that take years of schooling to master. While people have been producing and using artifacts forever (Donald, 1991; Geertz, 1973), we have little experience designing and teaching computational artifacts.

Artifacts play an absolutely central role in learning and understanding according to the philosophic roots that underlie the contemporary cognitive theories that Koschmann (1996; 1999) has identified as influential for theories of collaborative learning: situated action (Suchman, 1987), situated learning (Lave & Wenger, 1991), activity theory (Engeström et al., 1999), distributed cognition (Hutchins, 1996), dialogicality (Bakhtin, 1986), and critical inquiry (Dewey & Bentley, 1949/1991). This pivotal role of artifacts can be traced back to Hegel and Marx.

According to Hegel (1807/1967, p. 234 ff), the very basis of self-consciousness and sociality in mutual recognition is thoroughly mediated by the creation and use of artifacts – that embody human consciousness or meaning in their imposed form or design. Marx (1867/1976) argues that the production, circulation, and consumption of artifacts as commodities is both affected by the prevailing social relations and reproduces those relations – and influences how we understand and learn about contemporary artifacts; these commodities are essentially stored labor – physical and intellectual – that comes alive in use. Marx traces the social history of artifacts from simple tools through machinery to computational automated industry. For Husserl (1936/1989), meaning is established and historically sedimented in the form of artifacts; Heidegger (1927/1996) expands this analysis to argue that the life-world of our everyday involvements is structured as networks of meaningful artifacts. More recently, software is seen as a new form of stored meaning or intentionality (Keil-Slawik, 1992; Stahl, 1993a; Winograd & Flores, 1986). For instance, effects of “artificial intelligence” are accomplished by embedding human intelligence in software procedures and knowledge-bases.

Engelbart (1995) and Norman (1993) claim that it is artifacts that make us smart, by amplifying our very limited native abilities like short-term memory and attention. Others (Cole & Griffin, 1980; Pea, 1985) counter
that these artifacts change our tasks, rather than simply increasing our powers, but this still places artifacts centrally in our attempts to increase our intellectual capabilities. Donald (1991) argues that the entire enterprise of modern knowing and science only became possible with the development of artifacts like books, that provided external memories that could be circulated and that might outlive their creators. Papert (1980), reflecting on his own learning history, believes that playing with automobile gears as a young child “did more for my mathematical development than anything I was taught in elementary school. Gears, serving as models, carried many otherwise abstract ideas into my head” (p. vi).

If one looks closely at learning – from infancy to kindergarten, formal schooling, and on-the-job – one sees that artifacts (now including computational artifacts) are pervasive. While it is clear that a primary function of education (and socialization into culture generally) is to teach new-comers how to understand and use the available artifacts of one’s society or of its specialties, we have only narrow studies of how this takes place. For instance, Bruner (Bruner, 1990) discusses how children acquire the ability to follow and generate narratives as verbal cognitive artifacts, and Hall (Hall & Stevens, 1995) investigates how young students use design tools.

**How Artifacts are Understood**

In our pilot study of the use of a rocket simulation, it is clear that the process of coming to understand a computer simulation that models a scientific phenomenon is a complex process, that strains the cognitive abilities of middle school students. Without strong guidance from a teacher, the students would at best have treated the simulation as a video game, perhaps competing to get the highest rocket flight, but not investigating the scientific factors that might lead to success. Although students often make statements that sound like they understand how to construct certain kinds of knowledge, when one watches them struggling through the steps that are actually required one gains a much more detailed understanding of what is involved for a novice, what supports are helpful, and where problems typically arise. For instance, while the students in a pilot study we ran were proficient at taking averages of sets of numbers in a traditional math lesson, they ran into many problems when averaging their rocket simulation data. A major problem had to do with the organization of the data and of their averages on a data sheet. The two teams of students became confused about which rocket heights had been observed by which team, and which averages were associated with them.
While an adult experienced with scientific experiments can keep these things straight without thinking about it, the students had to learn this skill. They did this partially by negotiating with the teacher, who alerted them to problems and guided them back on track, and partially by collaboratively applying their own intellectual and communicative skills.

Our work and that of our current and past colleagues explores the use of gesture as well as language in understanding artifacts and in constructing shared understanding of artifacts. Our micro-ethnographic method (see below) is explicitly adapted to making learning visible by systematically attending to the sorts of gestures and bodily interactions that people use to co-construct the meaning of artifacts. In his seminal examples of micro-ethnographic analysis, Streeck (1983; 1993; 1996) focuses on the roles of gesture in making social understanding visible. LeBaron analyzes different forms of gesture that are successively used to build a shared vocabulary of meaningful gestural artifacts (LeBaron, 1998; LeBaron & Hopper, 1997; LeBaron & Koschmann, 1999; LeBaron & Koschmann, 2001; LeBaron & Streeck, 2000). Koschmann also highlights the role of gesture in educational settings (Koschmann et al., 1997; Koschmann & LeBaron, submitted; Koschmann, Ostwald & Stahl, 1998; Koschmann & Stahl, 1998).

Making Knowledge Building Visible

According to our theoretical framework, learning through interaction with artifacts is an inherently social process, involving either interaction with other people through the artifact or at least interacting with an artifact that was made by other people and that incorporates their intentions. For our research, collaborative interactions have an important characteristic: in order to collaborate, participants must make their ideas and their relationships visible to each other as part of their communication. That is, they make learning visible. As researchers, we can capture this in video or computer logs and analyze it. That way, we can see how students are relating to computational artifacts and what they are learning in the process. This overcomes the traditional problem of educational assessment, where it is assumed that learning is invisible to researchers and must be inferred from learning outcome measures. Thus, our approach avoids the restriction of educational assessment to the kinds of analyses of pre/post-test statistics and after-the-fact interviews that so often lead to “no significant difference” (Russell, 1999) results, which are of little value for software design purposes.

Of course, not all learning is made visible, so other methods to indirectly measure learning outcomes are necessary and complementary. But focusing
on the visible displays of learning prevents the common tendency to lose track of the learning in favor of secondary phenomena that seem easier to describe or quantify. For instance, much of the traditional literature on cooperative learning focuses on small group facilitation, rather than on cognitive and group learning processes. For a recent review of this literature, see (Brody & Davidson, 1998) reviewed by the PI (Stahl, 2000a). Even recent CSCL studies often miss the interesting learning phenomena, e.g., (Hakkarainen & Lipponen, in prep) and (Jong et al., in prep), reviewed by the PI (Stahl, in prep).

4. Micro-ethnographic Analysis of Interaction

Micro-ethnography is a rigorous social science that incorporates recent methods and findings from the analysis of verbal and visual human interaction. It uses digitized video to study interpersonal behavior at a detailed level. Micro-ethnography will be used in this project (a) to analyze the structure of negotiation in small group meetings in order to design software support of negotiation of ideas in personal and group Perspectives; (b) to study how people learn to use computer-based and internet-based cognitive artifacts that are part of a virtual biology laboratory; (c) to study specific distance-collaboration software as effective media for supporting perspectives in knowledge building.

Micro-Ethnography

For assessing software functionality and usability in this project, we adopt a recent tradition of human interaction analysis (Jordan & Henderson, 1995) that we refer to as “micro-ethnography.” This methodology builds on a convergence of conversation analysis (Sacks, 1992), ethnomethodology (Garfinkel, 1967), nonverbal communication (Birdwhistell, 1970), and context analysis (Kendon, 1990). An integration of these methods has only recently become feasible with the availability of videotaping and digitization that records human interactions and facilitates their detailed analysis. It involves close attention to the role that various micro-behaviors – such as turn-taking, participation structures, gaze, posture, gestures, and manipulation of artifacts – play in the tacit organization of interpersonal interactions. Utterances made in interaction are analyzed as to how they shape and are shaped by the mutually intelligible encounter as a holistic context – rather than being taken as expressions of individuals’ psychological intentions or of external social rules (Streeck, 1983).
Micro-ethnographic research typically involves the following components:

- A specific setting, or research site – such as several students gathered around a computer running VIRTUALBIOLOGYLAB.
- A detailed analysis of both audible and visible micro-behaviors, that are to be understood in terms of their embeddedness within the particular social and material environment – such as a biology class.
- A recognition that culture (which includes the meaning and use of shared artifacts) is a product and a process of naturally-occurring communication, simultaneously co-constructed and experienced by participants – and thereby made available for empirical study and interpretation by researchers.
- A use of recent technologies, like digitized video, that allow researchers to look at in detail the orderly performance of social life – such as the negotiation of learning between teacher and student or among collaborating peers.

Micro-ethnography can be adapted from the study of human-human interaction to that of human-computer interaction or computer-mediated collaboration. Our pilot studies suggest that such an adaptation of the methodology can be accomplished effectively.

**Micro-ethnography and Human-Computer Interaction**

Our research approach brings together software designers and micro-analytic researchers. We use micro-ethnography to analyze empirical student interactions with software artifacts. Techniques related to micro-ethnography, such as video analysis and conversation analysis, have previously been used to analyze human-computer interaction in limited cases (Bødker, 1989; Bødker, 1996; Frohlich & Luff, 1990; Hollan et al., 2000; McIlvenny, 1990; Nardi, 1996; Roschelle, 1996; Suchman, 1987; Suchman & Trigg, 1991). However, these cases typically did not analyze interactions at the micro-behavior level, including such things as gesture and posturing, that are important means of making understandings visible in face-to-face communication – suggestive exceptions include (Hutchins & Palen, 1998; Streeck, 1996). But, most importantly, these studies did not investigate learning technologies. Nor did they investigate learning taking place through the interactions. Those that did look at learning, like Roschelle (1996), did not use this to feed back into the design of the technology. Thus, our project is undertaking an approach that is unique in combining all three:
• Analysis of interaction at a micro level.
• Analysis of the learning taking place.
• Application of the analysis to revision of the technology.

Our own past work using micro-ethnography has begun to align this methodology with our project goals. Co-PI LeBaron (1998) shows through micro-ethnography how an architecture teacher goes through four stages of successive abstraction to define meaningful gestures, that the students then gradually adopt in their own presentations. By freezing key video frames and relating them to the speech and bodily behaviors of the teacher and students, LeBaron makes the teaching and learning process – that the participants are only tacitly aware of – visible to researchers. We also work with Koschmann, who has been engaged for almost 10 years in fine-grained studies of collaboration among medical students in a problem-based collaborative learning (PBL) curriculum (Glenn et al., 1999; Koschmann & Glenn, submitted; Koschmann et al., 1997; Koschmann et al., 2000; Koschmann, Ostwald & Stahl, 1998; Koschmann & Stahl, 1998; LeBaron & Koschmann, 2001). In particular, we have shown how group discussions raise learning issues for further study and how the status of these issues is negotiated by the students and a tutor. While we have investigated the role of a tutor in face-to-face PBL sessions, we have only recently begun to study the role of computer-based artifacts and media in distance-PBL sessions (Koschmann & LeBaron, submitted).

The proposed project will build upon the isolated pioneering efforts of ourselves and others, and put these methods together in a systematic way to apply them to the design of collaboration software.

Data Gathering and Analysis

Project staff, including Project Advisory Board members, will meet in workshops held monthly. The workshops not only review project progress and plan next steps, but they importantly include group data sessions for the analysis of data. The data gathering and analysis process for the VIRTUALBIOLOGYLAB trials will typically proceed through the following steps:

• Videotaping of students. Two or three students are gathered around a computer to interact asynchronously and remotely with other students. Cameras and microphones are set up to capture the facial expressions and body movements of all participants. The monitor image is also
captured. Microphones are arranged to capture all speech as clearly as possible and to distinguish the speakers.

- The video is combined (picture-in-picture) and time-code is burned in to provide a frame-by-frame reference system. A minute-by-minute record log is created, describing in a sentence or two what takes place each minute.

- A list of interesting episodes is created. Episodes are meaningful interactions lasting up to several minutes. Selected episodes are digitized and made available electronically. This allows them to be replayed easily, looped, freeze-frame, slowed down, and studied by project members at distant locations.

- A detailed transcript is created. It transcribes both speech and visible behaviors. Speech of different participants is color-coded. The transcripts are printed and posted on the Web with the digitized clips.

- Each episode is assigned to a project team member who “owns” that piece of data. The owner watches the clip many times to understand what is happening there.

- A data session is conducted at a group workshop. This is a collaborative analysis of the data’s empirical details. Usually, about two hours are spent on a single episode. The session is led by the owner of the data, who presents the episode and raises issues. The owner may audio-tape this session to preserve ideas and interpretations that come up.

The owner of the episode returns to a study of the video clip. At this point, the transcript may be revised and extended to include more details of interaction. The owner may invite other project team members to view and discuss the clip. The owner may present the clip at another data session. Finally, the owner drafts a micro-ethnographic analysis of the episode. This is distributed for comment. The analysis includes:

- A detailed description of the actions of all participants and their interactions.

- A discussion of what learning is evidenced in the data.

- A discussion of the role of any artifacts.

- A discussion of problems with the software, learning problems, etc.

The analyses of the episodes are reviewed by the project team, and various suggestions are made based on this:
• Proposed revisions to the software.
• Changes to the list of interesting episodes, such as the inclusion of additional episodes.
• Alterations to the research plan, such as scheduling additional usage sessions or changing the way they are conducted.
• Revisions to the research methodology and theoretical framework.

5. Study of a Virtual Biology Laboratory

The project will ultimately study collaborative knowledge building at a virtual biology laboratory used in geographically distributed high schools for advanced placement biology courses. A website containing the virtual biology lab is currently being developed at the University of Colorado; this project will assess its effectiveness in use by college freshmen and contribute to its iterative design (in Year I). The project will also develop a collaborative knowledge-building environment (in Year I) and integrate this with the lab (in Year II). Then (in Year III), micro-ethnographic methodology will be used to assess the use of this software – especially the use of computational Perspectives – by distributed high school students, and contribute to its iterative design.

Software Artifacts for a Virtual Biology Laboratory

The VIRTUALBIOLOGYLAB is a complete one-semester curriculum on the Web, intended to replace biology wet labs for college non-major freshmen and for high school advanced placement (AP) students. Each of 10 planned labs takes an estimated three hours for a student to work through – and enables students to conduct seminal experiments from the history of biology that would not be feasible in traditional physical wet labs (see attached letter of support from the developer). There are multiple kinds of artifacts composing the software: a guiding narrative, animations of lab equipment, simulations of lab procedures, data collection / analysis / graphing / display tools, background materials (theory, history, remedial text), links to related websites, and interactive assessment exercises. The virtual lab is designed to be used by students independent of any teacher guidance, although it is loosely coordinated with a biology theory course. Because biology AP students are sparsely scattered around a school district, it is convenient to have students conduct their labs on the Web.
Distance Negotiation

WEBPERSPECTIVES will be a software re-write of WEBGUIDE, a knowledge-building environment with Perspectives that the PI has been developing for several years (Stahl, 2000d). WEBGUIDE was always intended to have a negotiation component that would support the proposal, discussion, and decision by a group of users to promote a note from one Perspective to another (Stahl & Herrmann, 1999). Thus, a student could propose that a note from her personal Perspective be accepted by her team and be promoted to the team Perspective. After discussion and agreement by team members, the note would appear in the team Perspective. Similarly, notes could migrate all the way up the Perspective hierarchy to the class Perspective and become part of the knowledge accepted by the whole class. The planned negotiation mechanism has not yet been implemented because we are lacking an adequate understanding of how collaborative negotiation is conducted and how it should be supported.

We are currently videotaping meetings of a research group run by Clarence (Skip) Ellis. (Coincidentally, his group is designing collaboration software.) In Year I of this project, we will conduct a micro-ethnographic analysis of these tapes to study the structure of small group negotiation. We will then design and develop software functionality to support such processes within WEBPERSPECTIVES.

Semantic Relevance Agent for Intelligent Hyper-Linking

Research in KBEs like CSILE has shown that it is difficult to locate related ideas within a shared database of discussion notes (Hewitt et al., 1998; Hewitt & Teplovs, 1999). Therefore, in Year I, we will add functionality to WEBPERSPECTIVES to automatically locate the notes most closely related to a given note, such as a new idea just entered into my personal Perspective or an old note proposed for inclusion in a group Perspective.

We will use Latent Semantic Analysis (LSA) (Landauer & Dumais, 1997; Landauer et al., 1998; Stahl & dePaula, 2001) to analyze the semantic content of notes and to measure the semantic relatedness of pairs of notes. LSA is based upon a statistical analysis (singular value decomposition) of co-occurrences of terms in a large corpus of text. It determines the relatedness of words even if they did not occur together explicitly – hence the term “latent”. LSA incorporates some refinements that make its performance closer to that of humans than similar methods – see Discourse Processes (vol. 25, 1998) and Interactive Learning Environments (vol. 8, no. 2, 2000) for special issues of LSA assessment studies. The PI and his graduate assistant recently completed a four year project (sponsored by the
McDonnell Foundation CSEP Program) that successfully uses LSA in a Web-based educational system tested in middle school classrooms (Kintsch et al., 2000; Stahl & dePaula, 2001; Steinhart, 2000).

Automated linking of related notes will involve a fairly straight-forward application of LSA. It will be handled within the Perspectives server, running on a computer with access to the necessary files for LSA. A corpus of biology text (including the content of VIRTUALBIOLOGYLAB and the associated theory course) will be subjected to LSA analysis to define a semantic space. Periodically (e.g., each night) the site’s shared database of notes will be folded into this corpus to redefine the space and to compute the vector for each note within this space. In real time, when linking is requested for a new note, the new note’s vector can be quickly computed and a list of existing notes with the closest vectors in the semantic space can be produced without noticeable delay.

With this hyper-linking, students will be pointed from the themes of their own notes to places throughout the system and throughout the interactive knowledge-building discussions where the same and related themes occur. We will experiment with different interfaces to try alternative approaches to incorporating this functionality into KBEs. For instance, it can be left to users to ask for lists of notes related to a given note. Alternatively, an agent can automatically check to see if there are notes within a given closeness to certain notes: newly entered notes, notes proposed for negotiation, notes being read or edited, etc. The agent can then suggest that links be established from the given note to similar ones. The different interfaces can be tried out in our software trials.

An Open Architecture for Software Integration

A specific task of the proposed project is to structure the Perspective computation as a self-contained module with a well-defined application programming interface (API). This will form a Perspectives server, a Java application that runs on the Web server along with the database system. It will be separate from a client that runs in a Web browser on the client’s computer. This separation of functions into a server and a client will have many advantages. It will speed the functioning because the intensive computation of Perspective content will be done on a central server that is faster than typical student computers. Also, calls to the database system will take place locally rather than across the Internet. In terms of system development, it will mean that developers can build systems that incorporate Perspectives without having to worry about the Perspective algorithms or the database calls. They will use an API of the server that lets them request
the data that should be shown to a given user in a given Perspective. They can then just focus on how best to display this data in the interface client.

The Perspectives server will be a self-contained Java application. It will be released as Open Source with clear documentation on how to use it to get Perspective data for display. The data will be delivered as an XML text stream that can be used by any Web technology, such as HTML, PERL, or JAVA. Although it is anticipated that the Perspectives server will generally be used as a black box, its Open Source availability will allow programmers to modify it if necessary. We have already had requests from Germany and California for the release of such a Perspectives server.

The Perspectives server will be a form of middle-ware, operating between the database and the client software (see Figure 2). It will instantiate a three-tier, model-controller-view architecture that defines independent layers for the data schema or model, the data computation or control, and the interface display or view. The database management system can be any standard relational SQL system like MYSQL or ORACLE. The middle layer can be the Perspectives server or a stripped down version that does not compute Perspectives. And the interface can be any kind of applet, Web page, or Web application that conforms to the API standard.

6. Plan for Proposed Research

We recently submitted an NSF ROLE proposal to conduct research related to that proposed here. NSF’s staff review of our ROLE preproposal assigned it the highest possible ratings and suggested combining it with other NSF proposals. This CSS proposal and a related ITR proposal are in direct response to that suggestion. If both the ROLE and the CSS and/or ITR proposals are funded, then the grant budgets and scopes will be slightly
renegotiated to support a fully-staffed, year-round project. The ROLE proposal focuses on development of the micro-ethnographic approach for analyzing computer support of learning and collaboration; the ITR proposal focuses on development of the required technology; the current CSS proposal focuses on studying collaborative knowledge building achieved over distance and time with computer support. The three are closely complementary, although any one or two projects can be conducted independently. The following plan assumes that only this CSS proposal is funded.

The CSS project includes the design and development of WEBPERSPECTIVES, its integration with VIRTUALBIOLOGYLAB, and its eventual dissemination as Open Source software. Use of the software will be studied with small groups of college and high school students. Micro-ethnographic analyses will be conducted to understand how the software supports collaborative group knowledge building over distance and time. Assessment of these trials will feed back into the software development. Project findings will be broadly disseminated.

**Study of Collaboration**

In order to study computer-supported collaboration, college and high school students will be videotaped in pairs or triads working at computers. They will be engaging in team tasks that are part of VIRTUALBIOLOGYLAB. They will interact with other pairs or triads in their team or class through WEBPERSPECTIVES functions, negotiating content for group Perspectives. The interactions will be captured on videotape and computer logs.

The speech and gestures on videotape and the interactions with the software on the computer logs will provide our primary data, although some interviews will be used to triangulate our analyses. We will be particularly concerned with how groups of people using our software take advantage of certain functionality of the software to conduct their collaborative activities. Here are the kinds of issues we will be interested in focusing on in our micro-analyses:

- Are important issues of biology raised in the collaborative discussions? Are insights effectively shared and knowledge deepened?
- What collaborative knowledge-building activities can be identified in the data? How are they supported by the software? How are they hindered by the software?
• Do the software functionality, affordances, scaffolding, and support contribute to the user experience, do they further the collaborative effort, and do they motivate participation so that the features are used?

• How does learning of the software artifacts proceed so that users get over initial barriers and begin to take advantage of the support to accomplish things that could not previously be done?

• Do individuals, pairs, teams, and the class develop personal and group Perspectives containing different versions of the evolving shared knowledge? Does this organization into different Perspectives seem to help the collaborative knowledge building?

• Are negotiation mechanisms used to debate ideas from personal Perspectives and possibly promote them to team or class Perspectives? Do these mechanism seem to work intuitively? Do they help knowledge-building processes to converge on shared understanding?

• Does the semantic relevance agent provide useful suggestions of related notes? Are these suggestions used in future knowledge building?

The approach of micro-ethnography differs from the hypothesis-driven approach of many other social science methods. It requires an openness to the data. The history of ethnomethodology, conversation analysis, context analysis, and micro-ethnography suggests that this open attitude generally results in discoveries of important structures of social interaction. By designing our project around specific kinds of collaboration – with given software and specific curricular tasks, for instance – and by looking for episodes in the data that might shed light on the above guiding questions, we expect to gain a great deal of insight into how collaborative knowledge building is achieved by small groups over distance and time. We expect to see how computer support interacts with the collaborative activities and be able to iteratively modify our software to better support collaboration and knowledge building.

Cognitive artifacts like the computational Perspectives structure take time for people to learn how to use effectively and intuitively. We will be particularly interested in looking at episodes that reveal how this process takes place. At the beginning, there will have to be some form of instruction (to-do lists, seeded examples, explanations, etc.) in how to use Perspectives: where to enter tentative ideas, how to compare someone else’s ideas, when to propose the negotiation of an idea, and how to promote ideas to group Perspectives. Gradually, these activities will become natural and enter into the flow of collaborative exchanges. The guiding vision is that computational Perspectives will not only mirror or represent the personal
and group perspectives that form the structure of normal collaboration, but that they will allow people to manage this form of complexity well enough that collaboration will become more successful than ever as people master the technology. Micro-analysis will let us see if this in fact happens.

Project Schedule

Below is a timeline for major phases of the software development, the collaboration study, and the dissemination of findings:

<table>
<thead>
<tr>
<th></th>
<th>Software Development</th>
<th>Micro-ethnographic Assessment</th>
<th>Dissemination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall '01</td>
<td>Development of WebPerspectives</td>
<td>Study of group negotiation</td>
<td>Group 2001 in Boulder</td>
</tr>
<tr>
<td>Spring '02</td>
<td>Addition of negotiation</td>
<td>Pilot testing of VIRTUALBiologyLab</td>
<td>CSCL 2002 in Boulder</td>
</tr>
<tr>
<td>Summer '02</td>
<td>Addition of hyper-linking</td>
<td>Study of hyper-linking usage</td>
<td></td>
</tr>
<tr>
<td>Fall '02</td>
<td>Revision of architecture</td>
<td>Testing of VIRTUALBiologyLab</td>
<td>Conference demo</td>
</tr>
<tr>
<td>Spring '03</td>
<td>Integration of WebPerspectives with VIRTUALBiologyLab</td>
<td>Pilot testing of WebPerspectives with VIRTUALBiologyLab</td>
<td>European conference</td>
</tr>
<tr>
<td>Summer '03</td>
<td>Iterative revision</td>
<td>Study of Perspectives usage</td>
<td>Conference paper</td>
</tr>
<tr>
<td>Fall '03</td>
<td>Iterative revision</td>
<td>Testing of WebPerspectives with VIRTUALBiologyLab</td>
<td>Conference paper</td>
</tr>
<tr>
<td>Spring '04</td>
<td>System clean-up</td>
<td>Study of cognitive artifact usage</td>
<td>Journal article</td>
</tr>
<tr>
<td>Summer '04</td>
<td>Open Source version &amp; docs</td>
<td>Report on micro-analysis studies</td>
<td>Open Source release</td>
</tr>
</tbody>
</table>

Data collection and analysis issues including sampling and confidentiality will conform to rigorous research conventions and University of Colorado Human Subjects standards.

Project Assessment

We will engage in formative evaluation of our project throughout. That will be an important function of our Advisory Board, that includes assessment experts, and will form a regular part of monthly project workshops. We will check that we are making progress toward our project goals in accordance with the project timeline and are following our data analysis procedures.

The project’s micro-analytic approach provides a built-in assessment process. By videotaping sessions of students working with software artifacts, we will derive a formative evaluation of the artifacts. By the end of the project, we will be able to compare in a detailed and documented way how well our revised versions of collaboration software perform as compared to how they worked in the pilot studies and in earlier phases of the project.

In addition to the micro-ethnographic analysis that examines both how students learn with computer technologies and their learning processes as
revealed through their interactions (computer-mediated and face-to-face), it is important to understand how students relate to the technologies. In order to understand this, a triangulated approach to assessment will be adopted. Some students in the core trials of VIRTUALBIOLOGYLAB will be given a set of pre-assignment questions to gauge their prior knowledge and understanding of the concepts. Once they have completed the trial, they will be asked the same questions so that we can calculate their learning gains. In addition, we will interview these students in order to understand their perceptions of the artifacts as effective learning tools. This information will be gathered with each iteration and use of the software under development, and the comments and perceptions will be fed back into the development of software and the articulation of learning processes that involve computer software and computer-mediated collaboration. Understanding student perceptions of their experiences will also enable us to track our progress toward our research goals and to evaluate the effectiveness of the software by answering the critical question of, does it work: have we effectively supported distance collaborative knowledge building?

**Project Dissemination**

We will establish a website for both internal use and broad dissemination. The website will collect and coordinate materials and findings of the project. It will include logs of our videotapes, digitized clips of selected episodes, detailed transcripts, analyses of interactions, etc. It will also include all papers submitted to journals and conferences.

This project and its findings will be broadly disseminated in the CSCL, CSCW, HCI, education, and communication research communities through conferences and journals. It will be particularly prominent at CSCL 2002 and subsequent meetings of CSCL, AERA, CSCW, Group, ICLS, and WebNet.

It will also significantly impact the release of a published VIRTUALBIOLOGYLAB curriculum at the college and the high school level. VIRTUALBIOLOGYLAB will be published by commercial textbook publishers in both college and high school versions. This project will significantly improve the quality of this distance education curriculum by subjecting it to detailed assessment in use situations. The addition of WEBPERSPECTIVES to this curriculum will add an important collaborative dimension to it. WEBPERSPECTIVES will be made available as an Open Source component – the PI has already had requests from California and Germany for this.
Project Context and Resources

The proposed project will have considerable institutional support from the University of Colorado. The PI is a Research Professor active in both the Institute of Cognitive Science (ICS) and the Center for LifeLong Learning and Design (L3D). He is on the Steering Committee of the international ACM SIGGroup conference, GROUP ’01, and is Program Chair of the international CSCL ’02 conference (both being held in Colorado). These affiliations will provide support and visibility to this project.

Through its Advisory Board, the project brings together a valuable set of experienced professionals from education, communication, computer science, cognitive science, and project assessment.

7. Results from Previous Research

The project PIs and Advisory Board members (see Biographical Sketches) have conducted research projects (including a three-year CSS grant) and specific pilot projects that form a foundation for the proposed research. The PIs have previously developed knowledge-building environments with computational Perspectives for designers and students, have studied the theory of cognitive artifacts, and have engaged in micro-ethnographic analysis of students learning to use a scientific simulation. This and related work by others motivate new features and research issues. A previous Perspectives system will be extended with additional functionality and re-structured for integration with the virtual biology lab, for release as Open Source, for use in distance collaboration, and for micro-ethnographic analysis.

The proposed project builds upon a series of activities that we have already started to work on, and takes advantage of unique opportunities at the University of Colorado:

• The development of VIRTUALBIOLOGYLAB, that is currently underway in the Molecular, Cellular and Developmental Biology Department.

• The experience of the PI and his colleagues and students at the Center for LifeLong Learning and Design in the Computer Science Department in design and development of computer support for collaborative learning – including the Perspectives mechanism.
Proposals for Research

- The expertise of Advisory Board members and graduate students from the Communication Department in applying micro-ethnography to human interaction with technology.

- The participation of a number of people on the project Advisory Board, who bring complementary expertise in education, human-computer interaction, technology adoption, and project assessment.

Following are summaries of related work conducted by the PI and colleagues as preparation for this proposal and in related grants funded by NSF and other sources. Brief descriptions of Advisory Board members are given in the attached Biographical Sketches.

Pilot Studies

WEBGUIDE (http://www.cs.colorado.edu/~gerry/webguide)

WEBGUIDE is a knowledge-building environment for discussing topics via the Web developed by the PI and colleagues over the past two years. It has been used in a middle school environmental science class and in two college seminars on CSCL (Stahl, 1999c). WEBGUIDE goes beyond similar discussion-based systems by supporting the representation and development of personal and group Perspectives (Stahl, 1999a; Stahl, 1999b).

SIMROCKET (http://www.cs.colorado.edu/~gerry/previous/simrocket)

The PI designed and implemented SIMROCKET, a computer simulation of a rocket launch. The PI was invited to try it in a local Boulder school with five boys engaged in a model rocket science project. The teacher guided the students (grouped in front of two computers) to fire each of 7 virtual rockets with different characteristics six times and to average the resultant heights in order to predict the height of an 8th rocket. Project staff then engaged in micro-ethnographic analysis of this three-hour interaction – during data sessions in the Communication Department, the seminar on artifact theory, a summer workshop on micro-ethnography, and pilot sessions for this proposal (Stahl & Sanusi, 2001).

Artifacts Seminar (http://www.cs.colorado.edu/~gerry/readings)

The PI organized a seminar on artifact theory, primarily as a pilot project for this proposal. Core members of the project team met along with other faculty and graduate students from Communication, Education, Philosophy, and Computer Science. We reviewed theoretical texts on the nature of artifacts from cognitive science, CSCL, communication, cultural studies, psychology, philosophy, and social theory. We also held data sessions on
episodes from the SIMROCKET tapes. Out-of-class discussions were held in WEBGUIDE and we conducted a SIMROCKET experiment mediated by WEBGUIDE.

STATETHEESSENCE (http://www.cs.colorado.edu/~gerry/projects/essence)

STATETHEESSENCE is Web-based software developed by the PI to help middle school students develop their text summarization skills. It relied centrally on latent semantic analysis (LSA) technology, as developed by the co-PIs of the McDonnell Foundation grant, Walter Kintsch and Thomas Landauer. It was used in an interdisciplinary four-year research project at a local Boulder public school. After undergoing considerable revision and refinement based on testing with students, the software was shown to improve text summarization, particularly in cases where the original text was somewhat difficult for the student to understand (Kintsch et al., 2000; Stahl & dePaula, 2001; Steinhart, 2000). Evaluation of this software was conducted by means of controlled experiments and teacher ratings.

VIRTUALBIOLOGYLAB (http://www.virtuallaboratory.net)

VIRTUALBIOLOGYLAB is a Web-based curriculum to substitute for biology wet labs in introductory freshman biology courses for non-majors. Currently under development, it will soon consist of ten labs, each of which takes approximately three hours for a student to work through. Project staff has begun to review this software with the designer, Professor Mike Klymkowsky, who is closely involved in this project (see attached letter of support).

Grants Funded by NSF

Environmental Perspectives in a Middle School Classroom


This grant funded the initial implementation of WEBGUIDE as an integrated JAVA applet KBE supporting personal and group Perspectives. It was a joint effort between the PI, a middle school teacher, and a research group at the National Oceanographic and Atmospheric Administration (NOAA) labs in Boulder. WEBGUIDE was used by the students to collect notes on interviews and to formulate personal and team perspectives on a local environmental issue.

Organizational Memory and Organizational Learning
Proposals for Research

“Conceptual Frameworks and Computational Support for Organizational Memories and Organizational Learning (OMOL),” PIs: Gerhard Fischer, Gerry Stahl, Jonathan Ostwald, September 1997 – August 2000, $725,000, from NSF CSS Program #IRR-9711951.

This grant led to the current proposal’s focus on Web-based learning environments. It started with a model of computer support for organizations as Domain-Oriented Design Environments in which both domain knowledge and local knowledge are stored in the form of artifact designs and associated design rationale. This CSCW model evolved into one of Collaborative Information Environments, that emphasized the interactive, asynchronous, persistent discussion of concepts and issues within an organization. Gradually, interest in organizational learning aspects led to involvement in CSCL and the model of collaborative Knowledge-Building Environments (KBEs). A number of software prototypes were developed to explore the use of the Web as a communication and collaboration medium, including:

- **DYNACLASS**: A discussion forum for use in college courses. It features ties to an interactive glossary and bibliography, as well as email notification and specialized displays.
- **WEBGUIDE**: Differs from DYNACLASS in providing more control over management of notes; it features computational Perspectives.

Work on this grant led to the focus on KBEs as models of computer support for organizational memory and collaborative learning. In particular, it prototyped a number of different systems, each with useful functionality. As we tested and deployed these systems, we confronted serious issues of adoption and focused our concerns increasingly on socio-technical and social informatics issues: motivation, media competition, critical mass, social practices, seeding, management, re-seeding, convergence of ideas, peer-to-peer collaboration, deployment strategies. These issues led to a new research agenda (Stahl, 2000b), and ultimately to issues of this proposal. Results of WEBGUIDE trials were analyzed and presented at AERA, CSCL, ICLS, CILT, WebNet, and Group conferences (Stahl, 1999a; Stahl, 1999b; Stahl, 1999c; Stahl, 2000c; Stahl & Herrmann, 1999). These findings led to a recognition of the need for software architectures and components for KBEs as proposed in the current proposal.

**References**


Conference of the International Communication Association, San Francisco, CA.


Stahl, G. (in prep) Rediscovering CSCL. In T. Koschmann, R. Hall, & N. Miyake (Eds.), CSCL2: Carrying Forward the Conversation, Lawrence Erlbaum, Mahway, NJ.


Information Technology for Distributed Collaborative Learning in a Virtual Biology Lab

PROJECT SUMMARY

This project develops, deploys, and assesses innovative information technology to support distributed and collective learning. The technology, called WEBPERSPECTIVES, is integrated into VIRTUALBIOLAB, an online high school biology lab environment, to allow geographically distributed students to work in collaborative groups. WEBPERSPECTIVES is a server technology that organizes displays of information from a shared knowledge repository according to personal and group perspectives. The approach of computational perspectives is motivated by a theory of learning as collaborative knowledge building, in which the interaction of perspectives is seen as central to collaboration. The design and assessment of this technology within the virtual lab context takes advantage of micro-ethnographic video analysis of small groups of students using the software.

This research responds to the perceived potentials of distance education and collaborative learning. Distance education is increasingly seen as an attractive solution to several problems in contemporary schooling, including that of broader access to educational resources. At the same time, educational research emphasizes the importance of social interaction, communication, and collaboration for learning – something that is often overlooked in distance education environments. The WEBPERSPECTIVES technology is explicitly designed to support collaborative learning in distance education by representing the essential interaction among personal and group perspectives in collaborative knowledge building. The methodology of micro-ethnography analyzes this interaction in order to inform the design and assessment of such technology being used in naturalistic learning settings.

To students using VIRTUALBIOLAB, this Web-based system appears as a coherent educational environment consisting of a number of features. The centerpiece of each lab curriculum is a series of exercises that students must
work through, simulating the steps taken in conducting a specific biology experiment in a physical wet lab. Associated with these exercises are computer simulations, questions to be answered, background information pages, and related websites. In addition, there is a collaboration window consisting of a threaded discussion and various knowledge management tools. The threaded discussion displays all notes in a selected perspective. The perspective can be that of an individual student (the user or a colleague), that of a group of students who are working through the experiment together, that of the entire class which shares common tasks, or that of a particular knowledge management function like negotiation.

This will be the first attempt to use perspectives in distance education. The concept of computational perspectives derives from Vanevar Bush’s “trails” and Ted Nelson’s “transduction”; it was further developed in Hermes, the PI’s design-support hypertext system, and in WebGuide, his Web-based collaborative knowledge-building system. However, it has never before been integrated with a science learning environment or used in a distance education setting. While most of the components of this system have been developed in previous work by the PI and his colleagues, they will have to be integrated in the present project. This requires the development of a flexible architecture, modification of components to be compatible with the architecture, and development of new components to fill in missing functionality. Through iterative design and testing, computational perspectives will become capable of supporting collaboration in distance education.

Collaboration is always difficult and tricky. There are complex and subtle interactions and relationships to keep track of. A central question for research in this area is whether computational support of collaboration can be made intuitive enough that people using it will be able to work fluidly, enjoy it, and be motivated to participate fully. In order to investigate this, the project will videotape small groups of students using the system and analyze their experience at a very detailed interactional level using micro-ethnography. This is a rigorous qualitative method that can make collaborative learning visible to researchers. It can reveal: (a) how well students are understanding each other and successfully building knowledge collaboratively, (b) what and how the students are learning, (c) where the software design is producing problems.

This three-year research project will be closely associated with complementary research and development, including the iterative design of the virtual biology labs and the micro-ethnographic assessment of their use by college freshman at the University of Colorado. In its third year, the
The project will deploy the software (VIRTUALBIOLAB with WEBPERSPECTIVES) to Advanced Placement high school biology students in the Boulder Valley School District. The project will involve the PI—who has conducted computer science and interdisciplinary cognitive science research in many successful projects—and graduate students from Computer Science, Communication, and Education. There will also be a ten-person Advisory Board of people experienced in human-computer interaction, education research, interaction analysis, and virtual environment assessment.

PROJECT DESCRIPTION

1. The Potential of IT Collaboration Support in Distance Science Education

The Potential of Computer Support

Based on our own experiences with software in classrooms, we have found that computer-supported collaborative learning (CSCL) has a vast—and largely untapped—potential (Kintsch et al., 2000; Stahl, 2000c). Access to global sources of information is just one facet. In addition, computer simulations can transform conceptual representations into interactive worlds for inquiry. They can transcend real-world barriers of time, expense, geography, scale, expertise, etc. to allow students to engage with and experience phenomena that have until now been unapproachable—such as Nobel prize-winning biology experiments. Hypertext systems of information can personalize presentations to meet individual learning needs, providing links to both remedial and supplemental information. Communication media can promote collaboration in ways never before possible, as well as among people who could not hitherto interact, allowing students to work with other students with similar interests far away. Structured curricular databases and shared knowledge-building environments can support student learning processes by providing access to ideas of scientists and fellow students in persistent forms that can be thought about and inter-related. However, we have seen that students always use computer artifacts in ways not envisioned by the designers. So, careful study of the artifacts in naturalistic settings is critical to the development of effective educational technology. This project
Proposals for Research

will develop and integrate a VIRTUALBIOLAB Web-based system that embodies these potentials, and will assess the degree to which these potentials are achieved.

The Potential of Distance Learning

The general problems of computer-mediated education multiply substantially under the pressure to rush to distance learning. Universities and dot.com’s around the country have jumped on the distance education bandwagon, without necessarily thinking through the complex educational issues involved. It is true that distance education has the potential to address various pressing educational, social, geographic, and economic issues (Keegan, 1986). It is also true that the technical infrastructure that will enable this revolution in education to proceed is being quickly set in place. However, the design and development of the necessary tools, curriculum, and content lags far behind. In the dash to market, providers of distance education are likely to settle on software technologies that were developed for other uses and are inappropriate for educational applications, curricula that implement outmoded approaches like drill-and-practice, and content that has not been tested for its learning effects. We need to develop more new models of computer support for learning that are effective in distance learning. This is particularly true of support for collaboration in distance education approaches: they too often try to get by with generic conferencing technologies like chat windows. We have seen from extensive studies of experimental knowledge-building environments like CSILE/KNOWLEDGEFORUM (Scardamalia & Bereiter, 1996), KIE/WISE (Cuthbert, 1999), and COVIS (Pea, 1993) that collaboration support for learning can be powerful. Environments like these transcend rote learning of isolated facts by engaging groups of students in discussions and explorations of challenging and meaningful scientific issues. Our WEBPERSPECTIVES software is designed to be such an environment.

The Potential of Collaborative Learning

Our theory of learning (see next section) stresses the social nature of learning and recommends support for collaborative approaches to education. But we need more than generic chat windows to encourage students to engage in scientific inquiry through discussion with other students. To begin to achieve the potential of collaborative learning, we need specialized communication media that provide specific functionality and that can scaffold the exploration of leading scientific themes. Like anyone developing scientific understanding, students should have knowledge
management tools to organize, categorize, revise, summarize, question, and propose. If they are to learn how to approach a topic using scientific modes of thought that are new to them, then their interactions with other students should be scaffolded and guided through the computer interface. The integration of the WEBPERSPECTIVES collaboration medium with VIRTUALBIOLAB is designed to provide such support.

The Potential of Perspectives Support

Our theory claims that collaboration centrally involves interaction among multiple personal and group perspectives. Most communication, conferencing, and collaboration media provide no support for organizing contributions according to who made them, or for building group perspectives from selections out of personal perspectives. Some systems provide two fixed levels: personal and group – either by limiting access to various pieces of information or by defining personal and group workspaces. WEBPERSPECTIVES will be the first system that enables multiple levels of groups and subgroups, and allows new subgroups to be added interactively. Moreover, WEBPERSPECTIVES automatically includes information from certain perspectives in other perspectives (according to a user-defined “content inheritance” lattice of perspectives), so that all information accepted by a group is incorporated in the perspectives of that group’s subgroups and members. This provides a computationally supported knowledge management and knowledge-building approach appropriate to the structure of interpersonal collaboration.

2. Theory of Learning as Collaborative Knowledge Building

Learning as Knowledge Building

The term “learning” can refer to a wide range of phenomena, from the accumulation of facts to the development of deep understanding; from the experience of anything new to the mastery of specific complex skills. Attempts to assess learning range from assumptions that people are always learning to frustration that people cannot transfer what they have learned to new circumstances (Russell, 1999). For the assessment of learning in the sciences, we prefer the term “knowledge building” (Bereiter, 2000). This refers to the progressive construction of knowledge about a specific topic of inquiry. For instance, a group of students may gradually propose tentative answers to a scientific question, accumulate relevant data, debate alternative
arguments, and converge toward a deeper understanding of the phenomenon (Donald, 1991). This process of knowledge building takes place within a community of inquiry, in which experiments can be replicated, assumptions questioned, and insights discussed (Scardamalia & Bereiter, 1996). Knowledge building is an interpersonal process that can be observed and documented, unlike learning, which is generally taken to be a psychological phenomenon that can at best be inferred indirectly through tests or other outcomes (Lave, 1991).

The Collaborative Nature of Learning

Knowledge building is inherently collaborative. By “collaborate” we simply mean “to work together” (from the Latin com laborare). To build knowledge is to formulate theories and similar bodies of knowledge that are (or at least could, in principle be) shared by a group of people and that meet accepted criteria for contributions to knowledge (Latour & Woolgar, 1979). A high school biology class, for instance, must develop knowledge of biological phenomena and theories that meet criteria that gradually approach the standards of the field of biology. Although it may reside in the minds of individuals, knowledge is generally the result of interactions with other people, with cultural artifacts, and with shared language (Vygotsky, 1930/1978). It is often useful to conceptualize and analyze learning within “units of analysis” that include small groups, their tools, and their language (Engeström, 1999; Hutchins, 1996).

This project is guided by a theory of collaborative Knowledge-Building Environments (KBEs) that we are developing (Stahl, 2000b). This theory proposes the following principles:

- Collaborative knowledge building is a particular view of group learning that focuses on a range of activities that take place within communities, as opposed to focusing on learning as the transmission of bits of information to individual learners.

- Collaborative knowledge building takes place largely through the interaction among people with different understandings from multiple personal and group perspectives.

- Such knowledge building within groups can be helped by appropriately designed information technology (IT) that supports various knowledge-building activities and supports interaction among alternative perspectives.
In the following subsections, we discuss the model of knowledge building, the role of perspectives, and the potential of computer support.

**A Model of Collaborative Knowledge Building**

One approach to better understanding how to design computer support for collaborative knowledge building in social settings is to conceptualize the various constituent activities involved in individual and social knowledge building. Figure 1 from (Stahl, 2000b) provides a starting point for this, combining aspects of activity theory, situated learning, hermeneutic philosophy, and distributed cognition theory (Chaiklin & Lave, 1993; Cole, 1996; Engeström et al., 1999; Gadamer, 1960/1988; Hutchins, 1996; Lave & Wenger, 1991; Nardi, 1996).

The idea of this diagram is that knowledge building can proceed through many different activities. The sequential structure of the model is only illustrative of a paradigmatic conceptualization. We understand that these activities complexly overlap in practice. The possible relationships among the individual activities – and particularly the interactions between the personal and social – can be complex and varied. The purpose of the diagram is to suggest a number of distinct activities that could be supported by a KBE with multiple functionality; the sequential flow is not intended to imply a necessary order to the activities.

![Figure 1. A model of personal understanding and social knowledge building.](image)

A set of seminal books and articles in Computer-Supported Collaborative Learning (CSCL) has formulated a view of learning as a social process of collaborative knowledge building within communities of practice (Brown &
Proposals for Research

Campione, 1994; Brown & Duguid, 1991; Koschmann, 1996; Lave, 1991; Lave & Wenger, 1991; Pea, 1993; Scardamalia & Bereiter, 1996; Wenger, 1998). However, these texts do not make the set of cognitive and social activities that underlie such a view explicit in the manner attempted in our KBE theory. Starting in the lower left corner, Figure 1 shows a cycle of personal understanding. The rest of the diagram depicts how personal beliefs can be articulated in language and become part of social interaction. Note that the results of social knowledge building eventually feed into personal understanding, providing the evolving toolkit of culturally-based individual cognitive capabilities. The collaborative knowledge-building process begins with (a) the articulation in language and the confrontation of these statements (b) with alternatives from other perspectives (c). The interplay of perspectives proceeds through various interactional mechanisms, potentially culminating in the reduction of shared knowledge to a text or other persistent artifact (d-k).

Perspectives in Knowledge Building

According to the philosophy of interpretation (hermeneutics) human understanding is fundamentally perspectival. We construct knowledge from our situated perspective in the world: our historical position, cultural tools, and practical interests (Gadamer, 1960/1988; Habermas, 1981/1984; Heidegger, 1927/1996; Stahl, 1975). Computational support for knowledge building can represent our interpretive perspectives with computational Perspectives (Boland & Tenkasi, 1995; Nygaard & Sørgaard, 1987; Winograd & Flores, 1986). (In this proposal, Perspective—with-a•capital-P will refer to the proposed computational mechanism that mirrors human interpretive perspectives—with-a•lower-case-p.) In this sense, Knowledge-Building Environments (KBEs) with computational Perspectives are designed to support the essential structure of collaboration. A key working hypothesis of the proposed project is that KBEs benefit from an approach that represents the perspectival nature of collaboration. A goal of the research is to facilitate the incorporation of a computational Perspectives mechanism in KBEs for distance education.

Computational Perspectives have been explored by the PI in a number of software prototypes, in his dissertation system, and in his theoretical publications (Stahl, 1993a; 1993b; 1995; 1998; Stahl & Herrmann, 1999; Stahl et al., 1995). In a single-user system, computational Perspectives may correspond to different domains or professional viewpoints on a design
problem, such as electrical, plumbing, structural, and heating concerns in architecture (Fischer et al., 1993a; Fischer et al., 1993b). In a KBE to support collaboration, computational Perspectives typically provide personal or group workspaces for the development of different sets of ideas. In this way, they can model the relationships among the various personal and group interpretive perspectives at work in the construction of collaborative knowledge. This project will introduce computational Perspectives into distance education.

**Computer Support for Collaborative Knowledge Building**

The form of IT that we are interested in – collaborative Knowledge-Building Environments – represents a distinctive approach that overlaps related work in Educational Technology, Computer-Supported Collaborative Learning (CSCL), and Computer-Supported Cooperative Work (CSCW). IT support for learning has traditionally been oriented toward the transmission of information to individual students. Even where it is based on a view of student construction of knowledge, as with Intelligent Tutoring Systems (ITS) for algebra or physics, the goal is measured by testing the incorporation of pre-defined content or methods into the individual’s understanding (Wenger, 1987). A more student-centered, constructivist approach is taken by Interactive Learning Environments (ILE), which might, for instance, allow students to create ecologies in SIMLIFE to learn biology, or programs in Turtle LOGO to explore math concepts (Papert, 1980). In contrast, a KBE primarily supports the group process and leaves matters of content up to the participants (which may include a teacher who raises particular content issues and helps maintain focus, or a website with content and scaffolding). In this way, it applies CSCW approaches to CSCL.

A review of CSCW technology for groups (Kraemer & Pinsonneault, 1990) distinguishes group communication support systems (GCSSs) from decision support (GDSSs). GCSSs are specific communication media like email, chat, threaded discussion, and video-conferencing. In providing computational tools for group decision making, GDSSs add tools for specific types of group interactions (e.g., voting), but tend to support isolated, focused activities that collate the work products and opinions of individual members of a group. In contrast, a KBE aims to support a broad spectrum of knowledge-building activities – such as activities (a) to (k) in our model – in a way that allows deep knowledge to evolve and emerge over time. It supports the construction and interaction of alternative formulations of knowledge. It also supports the interplay of individuals and groups more comprehensively, through integrated mechanisms of divergent
computational Perspectives and convergent negotiation processes that treat the group as more than just the sum of the individuals.

Assessments of CSCL and CSCW systems have defined a number of key issues for evaluating the problems and successes of such systems. For instance, in simple threaded discussion forums common problems include: short threads (a tendency for discussions to die quickly), low participation (lack of motivation to participate), few cross-references (little convergence of ideas), and superficial content (minimal depth of investigation) (dePaula, 1998; Guzdial & Turns, 2000; Hewitt & Teplovs, 1999). On the other hand, GDSSs and GCSSs attempt to decrease communication barriers within the group, while increasing task-oriented focus, depth of analysis, and decision quality (Connolly, 1997; Kraemer & Pinsonneault, 1990). Social informatics studies have raised additional issues of software deployment and adoption in addition to questions of usability and utility (Kling, 1999). These are some of the dimensions along which KBEs must be assessed within realistic learning and working social contexts.

In summary, we want to design, develop, and assess KBEs that go beyond generic chat and discussion systems (that tend to encourage exchange of personal opinions or isolated facts, but not deep shared understanding and critical inquiry). Such systems should include specific tools and structures to promote on-going debate, knowledge management, and group decision-making; however, knowledge building should go beyond the management and dissemination of existing knowledge to support the emergence of qualitatively new, increasingly shared knowledge within a community (Engelbart, 1962; Engelbart, 1995). The following section discusses tools to support this within the proposed project.

3. IT Artifacts for Supporting Collaboration in a Virtual Biology Lab

IT Artifacts for Biology Lab

The VIRTUALBIOLOGYLAB is a complex network of interdependent artifacts. It is a complete one-semester curriculum on the Web, intended to replace biology wet labs for high school advanced placement (AP) students. Each of 10 planned labs takes an estimated three hours for a student to work through – and enables students to conduct seminal experiments from the history of biology that would not be feasible in traditional physical wet labs (see attached letter of support from the developer). There are multiple kinds of artifacts composing the software: a guiding narrative, animations of lab
equipment, simulations of lab procedures, data collection / analysis / graphing / display tools, background materials (theory, history, remedial text), links to related websites, and interactive assessment exercises. The virtual lab is designed to be used by students independent of any teacher guidance, although it is loosely coordinated with a biology theory course. Because biology AP students are sparsely scattered around a school district, it is convenient to have students conduct their labs on the Web.

**IT Artifacts for Collaborative Knowledge Building**

WEBPERSPECTIVES is a knowledge-building environment to support collaborative learning. It provides a collaboratively constructed and shared external memory medium on the Web. The display is dynamically computed to show a hierarchy of notes arranged as a personal or group Perspective on the persistent, asynchronous discussion. This Perspective mechanism is an artifact that people must learn how to use and navigate to mirror and support the interpersonal relationships of collaboration. WEBPERSPECTIVES also incorporates a variety of knowledge management functions that must be learned in order to manipulate the ideas stored in the system and to build effective shared knowledge. Most of the functionality of WEBPERSPECTIVES has already been pilot tested in WEBGUIDE (Stahl, 2000c) and will not be further described in this proposal. Only the central Perspectives mechanism, which must be adapted and integrated in this project, and a proposed new search and hyper-linking tool will be described.

**IT Artifacts for Perspectives**

The core of WEBPERSPECTIVES is a Perspectives server that queries a database of textual contributions to the on-going discussion and provides user interface clients with the notes that are to be displayed in the Perspective requested by the user. The Perspectives associated with a particular knowledge-base form a non-cyclical lattice (each Perspective may have multiple parents or super-groups and multiple children or subgroups). The knowledge-base for a biology class would typically have a Perspective for questions and ideas shared by the whole class, several subgroup Perspectives for teams of students who work together, a personal Perspective for each student, and comparison Perspectives that bring together contributions from the Perspectives of all members of a group (see figure 2). The algorithm for computing the contents of a selected Perspective is outlined in figure 3. This algorithm allows new perspectives to be defined dynamically and interactively through entries in the knowledge-base.
Figure 2. A sample diagram showing the inheritance of content among Perspectives for a biology class. The class is divided into small work groups: Team A, B, C, D. Some students belong to special Projects I or II. Note how S4, a student’s personal perspective, inherits information from the class perspective, Team B, and Project I, and how notes in S4 are in turn inherited in the corresponding comparison perspectives.
visible (p) =

for all nodes, if node was created in p, add node to result set

for each parent of p in the perspective lattice, add the result set of visible(parent)

for each node deleted in p, remove node from result set

for each node edited in p, remove node from result set

for each node virtually copied into p, add node to result set

return ordered set of nodes in the result set

Figure 3. Recursive algorithm to compute set of nodes visible in perspective p.

Computational Perspectives provide a new, dynamic, personalized form of on-line information management (Stahl, 1995). A Perspective defines an electronic workspace in which a person or group can develop ideas and manage information that belongs together – for instance because it represents the beliefs and viewpoint of a particular person, group, domain, or intellectual position. Perspectives structure a shared information space so that special coherent views can be built up and displayed. The mechanism of computational Perspectives is very general and flexible.

The design philosophy behind computational Perspectives as implemented in WEBPERSPECTIVES is that users have complete control over the content in their personal Perspectives. Thus, if my personal Perspective inherits conflicting ideas from different team Perspectives that I belong to, I can delete, edit, and rearrange those ideas at will. Other users can view the contents of my personal Perspective (except for content that I have designated as private) and they can copy items, link to them, initiate public discussions of them, and propose them for incorporation in team Perspectives – but none of this affects how the content of my Perspective is displayed to me. This allows me to build my own Perspective on the topics that are under consideration by the group. I can see what knowledge others are building, incorporate that knowledge into my Perspective, or join in with others to share, discuss, and negotiate. The same design philosophy applies of course to team Perspectives: team members jointly (through negotiation processes) have complete control over the content of their team Perspective.

Inheritance is a central defining mechanism of computational Perspectives as used in this proposal. The ability to define arbitrarily complex networks of Perspectives with multiple layers of sub-groups between the group Perspective and the individual personal Perspectives, and to have the automatic inheritance of content through the network distinguishes this
approach from all other systems of “views” and “perspectives.” Inheritance in this sense is not class inheritance, but “content inheritance.” A given Perspective can inherit content from multiple other Perspectives. This content is aggregated (logical union) in the given Perspective, where it can be over-ridden with edits, deletions, rearrangements, virtual-copying (linking), and additions. The inheritance mechanism is derived from efficient approaches explored in hypermedia, including “delta memory” and “transclusion” (Boborow & Goldstein, 1980; McCall et al., 1990; Mittal et al., 1986; Nelson, 1981; Nelson, 1995). For a discussion of related work, see (Stahl & Herrmann, 1999).

Because new Perspectives can be defined (either in advance or during system use) to inherit from any (non-cyclical) other Perspectives, it is generally useful to define “comparison Perspectives” that aggregate the ideas from team members, including those ideas that have not been agreed upon and migrated to the team Perspective. This is handy for keeping an eye on what one’s fellow team members are thinking. Typically, we have set up the inheritance network of Perspectives to have a diamond-shaped profile, diverging out from the total group Perspective via teams to all the personal Perspectives, and then converging back via team comparisons to the group comparison Perspective. This models a collaborative knowledge-building process that combines divergent brainstorming and convergent negotiation. Various special displays can also be computed using this inheritance computation by treating discussions, negotiations, historical archives, etc. as pseudo-Perspectives that have special inheritance and exclusion rules. Once the Perspectival data computation has been returned, the content can be displayed in specialized interfaces that provide different kinds of functionality useful for further knowledge building.

**IT Artifacts for Intelligent Hyper-Linking**

Research in KBEs like CSILE has shown that it is difficult to locate related ideas within a shared database of discussion notes (Hewitt et al., 1998; Hewitt & Teplovs, 1999). Therefore, in Year III, we will add functionality to automatically locate the notes most closely related to a given note, such as a new idea just entered into my personal Perspective or an old note proposed for inclusion in the group Perspective. We will use Latent Semantic Analysis (LSA) (Landauer & Dumais, 1997; Landauer et al., 1998; Stahl & dePaula, 2001) to analyze the semantic content of notes and to measure the semantic relatedness of pairs of notes. LSA is based upon a statistical analysis (singular value decomposition) of co-occurrences of terms in a large corpus of text. It determines the
relatedness of words even if they did not occur together explicitly – hence the term “latent”. LSA incorporates some refinements that make its performance closer to that of humans than similar methods (see special issues of Discourse Analysis 1997 and Interactive Learning Environments 2000 on LSA assessment studies). The PI and his graduate assistant are currently completing a four year project (sponsored by the McDonnell Foundation CSEP Program) that successfully uses LSA in a Web-based educational system tested in middle school classrooms (Kintsch et al., 2000; Stahl & dePaula, 2001; Steinhart, 2000).

Automated linking of related notes will involve a fairly straight-forward application of LSA. It will be handled within the Perspectives Server, running on a computer with access to the necessary files for LSA. A corpus of biology text (including the content of VIRTUALBIOLAB and the associated theory course) will be subjected to LSA analysis to define a semantic space. Periodically (e.g., each night) the site’s shared database of notes will be folded into this corpus to redefine the space and to compute the vector for each note within this space. In real time, when linking is requested for a new note, the new note’s vector can be quickly computed and a list of existing notes with the closest vectors in the semantic space can be produced without noticeable delay.

With this hyper-linking, students will be pointed from the themes of their own notes to places throughout the system and throughout the interactive knowledge-building discussions where the same and related themes occur. We will experiment with different interfaces to try alternative approaches to incorporating this functionality into KBEs. For instance, it can be left to users to ask for lists of notes related to a given note. Alternatively, a software agent can automatically check to see if there are notes within a given closeness to certain notes: newly entered notes, notes proposed for negotiation, notes being read or edited, etc. The agent can then suggest that links be established from the given note to similar ones. The different interfaces can be tried out in our trials.

**An Architecture to Integrate the IT Artifacts**

A specific task of the proposed project is to structure the Perspective computation as a self-contained module with a well-defined application programming interface (API). This will form a Perspectives Server, a Java application that runs on the Web server along with the database system. It will be separate from a client that runs in a Web browser on the client’s computer. This separation of functions into a server and a client will have many advantages. It will speed the functioning because the intensive
computation of Perspective content will be done on a central server that is faster than typical student computers. Also, calls to the database system will take place locally rather than across the Internet. In terms of system development, it will mean that developers can build systems that incorporate Perspectives without having to worry about the Perspective algorithms or the database calls. They will use an API that lets them request data that should be shown to a given user in a given Perspective. They can then just focus on how best to display this data in the interface.

The Perspectives Server will be a self-contained Java application. It will be released as Open Source with clear documentation on how to use it to get Perspective data for display. The data will be delivered as an XML text stream that can be used by any Web technology, such as HTML, Perl, or Java. Although it is anticipated that the Perspectives Server will generally be used as a black box, its Open Source availability will allow programmers to modify it if necessary.

The Perspectives Server will be a form of middle-ware, operating between the database and the client software (see Figure 4). It will instantiate a three-tier, model-controller-view architecture that defines independent layers for the data schema or model, the data computation or control, and the interface display or view. The database management system can be any standard relational SQL system like mySQL or Oracle. The middle layer can be the Perspectives Server or a stripped down version that does not compute Perspectives. And the interface can be any kind of applet, Web page, or Web application that conforms to the API standard (see Figure 4).
4. Methodology for Assessing IT Artifacts

Micro-Ethnography

For assessing software functionality and usability in this project, we adopt a recent tradition of human interaction analysis (Jordan & Henderson, 1995) that we refer to as “micro-ethnography.” This methodology builds on a convergence of conversation analysis (Sacks, 1992), ethnomethodology (Garfinkel, 1967), nonverbal communication (Birdwhistell, 1970), and context analysis (Kendon, 1990). An integration of these methods has only recently become feasible with the availability of videotaping and digitization that records human interactions and facilitates their detailed analysis. It involves close attention to the role that various micro-behaviors – such as turn-taking, participation structures, gaze, posture, gestures, and manipulation of artifacts – play in the tacit organization of interpersonal interactions. Utterances made in interaction are analyzed as to how they shape and are shaped by the mutually intelligible encounter as a holistic context – rather than being taken as expressions of individuals’ psychological intentions or of external social rules (Streeck, 1983).

Micro-ethnographic research typically involves the following components:

- A specific setting, or research site – such as several students gathered around a computer running VIRTUALBIOLAB.
- A detailed analysis of both audible and visible micro-behaviors, which are to be understood in terms of their embeddedness within the particular social and material environment – such as a biology class.
- A recognition that culture (which includes the meaning and use of shared artifacts) is a product and a process of naturally-occurring communication, simultaneously co-constructed and experienced by participants – and thereby made available for empirical study and interpretation by researchers.
- A use of recent technologies, like digitized video, that allow researchers to look at in detail the orderly performance of social life – such as the negotiation of learning between teacher and student or among collaborating peers.

Micro-ethnography can be adapted from the study of human-human interaction to that of human-computer interaction or computer-mediated collaboration. Our pilot studies suggest that such an adaptation of the methodology can be accomplished effectively.
Micro-ethnography and Human-Computer Interaction

Our research approach brings together educational software design and micro-analytic research. We use micro-ethnography to analyze empirical student interactions with educational software artifacts. Techniques related to micro-ethnography, such as video analysis and conversation analysis, have previously been used to analyze human-computer interaction in limited cases (Bødker, 1989; Bødker, 1996; Frohlich & Luff, 1990; Hollan et al., 2000; McIlvenny, 1990; Nardi, 1996; Roschelle, 1996; Suchman, 1987; Suchman & Trigg, 1991). However, these cases typically did not analyze interactions at the micro-behavior level, including such things as gesture and posturing, which are important means of making understandings visible in face-to-face communication (suggestive exceptions include (Hutchins & Palen, 1998; Streeck, 1996)). But, most importantly, these studies did not investigate learning technologies. Nor did they investigate learning taking place through the interactions. Those that did look at learning (like Roschelle (1996)) did not use this to feed back into the design of the technology. Thus, our project is undertaking an approach that is unique in combining all three:

• Analysis of interaction at a micro level.
• Analysis of the learning taking place.
• Application of the analysis to revision of the technology.

The Assessment Process

Our gathering and analysis of data involves the PI working closely with the graduate student project members. In addition, our Advisory Board members participate in workshops held monthly. The workshops not only review project progress and plan next steps, but they importantly include group data sessions for the analysis of data. The data gathering and analysis process for the VIRTUALBIOLAB trials will typically proceed through the following steps:

• Videotaping of students. Two or three students are gathered around a computer. Cameras and microphones are set up to capture the facial expressions and body movements of all participants. The monitor image is also captured. Microphones are arranged to capture all speech as clearly as possible and to distinguish the speakers.
• The video is combined (picture-in-picture) and time-code is burned in to provide a frame-by-frame reference system.
• A minute-by-minute record log is created, describing in a sentence or two what takes place each minute. This is typically done by a graduate student and reviewed by a PI. The log may be revised later.

• A list of interesting episodes is created. Episodes are meaningful interactions lasting up to several minutes. The list is discussed by the whole project team at a group workshop.

• Selected episodes are digitized and made available electronically. This allows them to be replayed easily, looped, freeze-framed, slowed down, and studied by project members at distant locations.

• A detailed transcript is created. It transcribes both speech and visible behaviors. Speech of different participants is color-coded. The transcripts are printed and posted on the Web with the digitized clips.

• Each episode is assigned to a project team member who “owns” that piece of data. The owner watches the clip many times to understand what is happening there.

• A data session is conducted with the whole project team at a group workshop. This is a collaborative analysis of the data’s empirical details. Usually, about two hours are spent on a single episode. The session is led by the owner of the data, who presents the episode and raises issues. The owner may audio-tape this session to preserve ideas and interpretations that come up.

The owner of the episode returns to a study of the video clip. At this point, the transcript may be revised and extended to include more details of interaction. The owner may invite other project team members to view and discuss the clip. The owner may present the clip at another data session. Finally, the owner drafts a micro-ethnographic analysis of the episode. This is distributed for comment. The analysis includes:

• A detailed description of the actions of all participants and their interactions.

• A discussion of what learning is evidenced in the data.

• A discussion of the role of any artifacts.

• A discussion of problems with the software, learning problems, etc.

The analyses of the episodes are reviewed by the whole project team and various suggestions are made based on this:

• Proposed revisions to the software.
• Changes to the list of interesting episodes, such as the inclusion of additional episodes.

• Alterations to the research plan, such as scheduling additional usage sessions or changing the way they are conducted.

• Revisions to the research methodology and theoretical framework.

5. Results from Previous Research

The proposed project builds upon a series of activities that we have already started to work on. These activities – that either grew out of our previous engagements or were conducted to explore the basis for this proposal – have led to the design of our project and are suggestive of its likely success. The project takes advantage of unique opportunities at the University of Colorado:

• The development of VIRTUALBIOLAB, which is currently underway in the Molecular, Cellular and Developmental Biology Department.

• The experience of the PI and his colleagues and students at the Center for LifeLong Learning and Design in the Computer Science Department in design and development of computer support for collaborative learning – including the Perspectives mechanism.

• The expertise of Advisory Board members and graduate students from the Communication Department in applying micro-ethnography to the human interaction with technology.

• The participation of a number of people on the project Advisory Board, who bring complementary expertise in education, human-computer interaction, technology adoption, and project assessment.

Following are summaries of related work conducted by the PI and colleagues as preparation for this proposal and in related grants funded by NSF and other sources:

Pilot Studies

VIRTUALBIOLOGYLAB (http://www.virtuallaboratory.net)

VIRTUALBIOLOGYLAB is a Web-based curriculum to substitute for biology wet labs in introductory freshman biology courses for non-majors.
Currently under development, it will soon consist of ten labs, each of which takes approximately three hours for a student to work through. Project staff has begun to review this software with the designer, Professor Mike Klymkowsky, who is closely involved in this project (see attached letter of support).

**WEBGUIDE** ([http://www.cs.colorado.edu/~gerry/webguide](http://www.cs.colorado.edu/~gerry/webguide))

WEBGUIDE is a knowledge-building environment for discussing topics via the Web developed by the PI and colleagues over the past two years. It has been used in a middle school environmental science class and in two college seminars on CSCL (Stahl, 1999d). WEBGUIDE goes beyond similar discussion-based systems by supporting the representation and development of personal and group Perspectives (Stahl, 1999b; Stahl, 1999c).

**SIMROCKET** ([http://www.cs.colorado.edu/~gerry/previous/simrocket](http://www.cs.colorado.edu/~gerry/previous/simrocket))

The PI designed and implemented SIMROCKET, a computer simulation of a rocket launch. The PI was invited to try it in a local Boulder school with five boys engaged in a model rocket science project. The teacher guided the students (grouped in front of two computers) to fire each of 7 virtual rockets with different characteristics six times and to average the resultant heights in order to predict the height of an 8th rocket. Project staff then engaged in micro-ethnographic analysis of this three-hour interaction – during data sessions in the Communication Department, the seminar on artifact theory, a summer workshop on micro-ethnography, and pilot sessions for this proposal.

**Artifacts Seminar** ([http://www.cs.colorado.edu/~gerry/readings](http://www.cs.colorado.edu/~gerry/readings))

The PI organized a seminar on artifact theory, primarily as a pilot project for this proposal. Core members of the project team met along with other faculty and graduate students from Communication, Education, Philosophy, and Computer Science. We reviewed theoretical texts on the nature of artifacts from cognitive science, CSCL, communication, cultural studies, psychology, philosophy, and social theory. We also held data sessions on episodes from the SimRocket tapes. Out-of-class discussions were held in WebGuide and we conducted a SIMROCKET experiment mediated by WEBGUIDE.

**STATETHEESSENCE** ([http://www.cs.colorado.edu/~gerry/projects/essence](http://www.cs.colorado.edu/~gerry/projects/essence))

STATETHEESSENCE is Web-based software developed by the PI to help middle school students develop their text summarization skills. It relied centrally on latent semantic analysis (LSA) technology, as developed by co-PIs Walter Kintsch and Thomas Landauer. It was used in an
interdisciplinary four-year research project at a local Boulder public school. After undergoing considerable revision and refinement based on testing with students, the software was shown to improve text summarization, particularly in cases where the original text was somewhat difficult for the student to understand (Kintsch et al., 2000; Stahl & dePaula, 2001; Steinhart, 2000). Evaluation of this software was conducted by means of controlled experiments and teacher ratings.

**Grants Funded by NSF**

*Organizational Memory and Organizational Learning*

“Conceptual Frameworks and Computational Support for Organizational Memories and Organizational Learning (OMOL),” PIs: Gerhard Fischer, Gerry Stahl, Jonathan Ostwald, September 1997 – August 2000, $725,000, from NSF CSS Program #IRR-9711951.

This grant led to the current proposal’s focus on Web-based learning environments. The OMOL project started from a model of computer support for organizations as Domain-Oriented Design Environments (DODEs) in which both domain knowledge and local knowledge are stored in the form of artifact designs and associated design rationale. This CSCW model evolved into one of Collaborative Information Environments (CIEs), that emphasized the interactive, asynchronous, persistent discussion of concepts and issues within an organization. Gradually, interest in organizational learning aspects led to involvement in CSCL and the model of collaborative Knowledge-Building Environments (KBEs). A number of software prototypes were developed to explore the use of the Web as a communication and collaboration medium, including:

- **DYNACLASS**: A discussion forum for use in college courses. It features ties to an interactive glossary and bibliography, as well as email notification and specialized displays.

- **WEBGUIDE**: Differs from DYNACLASS in providing more control over rearrangement of notes; features computational Perspectives.

Work on this grant led to the focus on KBEs as models of computer support for organizational memory and collaborative learning. In particular, it provided a number of different systems, each with useful functionality. As we tested and deployed these systems, we confronted serious issues of adoption and focused our concerns increasingly on socio-technical and social informatics issues: motivation, media competition, critical mass, social practices, seeding, management, re-seeding, convergence of ideas,
peer-to-peer collaboration, deployment strategies. These issues led to a new research agenda (Stahl, 2000a) and this proposal.

Environmental Perspectives in a Middle School Classroom


This grant funded the initial implementation of WEBGUIDE as an integrated JAVA applet KBE supporting personal and group Perspectives. It was a joint effort between the PI, a middle school teacher, and a research group at the National Oceanographic and Atmospheric Administration (NOAA) labs in Boulder. The teacher taught an environmental science class in which he wanted to spend the year having his students interview various adults and construct a set of contrasting perspectives (conservationist, regulatory, business, community) on a particular local environmental issue that the students had previously been involved in. WEBGUIDE was used by the students to collect notes on their interviews and to formulate personal and team perspectives on the issue. Results of this software trial were analyzed and presented at the AERA, CSCL, ICLS, CILT, WebNet, and Group conferences (Stahl, 1998b; Stahl, 1999a; Stahl, 1999b; Stahl, 1999c; Stahl, 1999d; Stahl, 2000b; Stahl & Herrmann, 1999). These findings led to a recognition of the need for software architectures, standards, and components for KBEs as proposed in the current proposal.

Interoperability among Knowledge-Building Environments

“Interoperability Among Knowledge-Building Environments,” PI: Gerry Stahl, September 1999 – August 2000, $9,124, from NSF-funded Center for Innovative Learning Technology (CILT), Subcontract #17-000359 under NSF grant #EIA-9720384.

This was a seed grant whose purpose was to stimulate collaboration among KBE research groups. This grant resulted in a semester-long student project involving three graduate and three undergraduate students creating an XML DTD that defines a data format for data imported from several different KBE prototypes and displayed in a Web browser using XSL. The grant supported a workshop organized by the PI, entitled “Collaborating on the Design and Assessment of KBEs in the 2000's” at CSCL '99 at Stanford. This workshop attracted over 60 participants and was preceded by an on-line discussion of 28 submitted position papers. This grant led to an emphasis on collaboration among KBE research groups and the need to develop and disseminate theory and methodology for developing educational artifacts, as proposed here.
Other Grants

Incorporating Automated Text Evaluation into Collaborative Software Environments

“Allowing Learners to be Articulate: Incorporating Automated Text Evaluation into Collaborative Software Environments,” PIs: Walter Kintsch, Gerhard Fischer, Thomas Landauer (Stahl served as co-PI and primary software developer), calendar 1997-2000, $1,400,000 for four-years, from the McDonnell Foundation’s Cognitive Science in Education (CSEP) program.

This grant supported the design, implementation, and testing of STATETHEESSENCE, Web-based software to teach middle school students summarization skills.

New Media to Support Collaborative Knowledge Building


The grant was to test WEBGUIDE in the Artifacts Seminar and to make a number of technical improvements to WEBGUIDE’S functionality.

6. Plan for Proposed Research

We recently submitted an NSF ROLE proposal to conduct research related to that proposed here. NSF’s staff review of our ROLE preproposal assigned it the highest possible ratings and suggested combining it with an ITR proposal. This ITR proposal is a direct response to that suggestion. If both the ROLE and the ITR proposals are funded, then the ITR will extend the PI’s support to 12 months, allowing him to direct this research full-time. It will also support 3 graduate students to focus on the technology development for the distance education component of the ROLE proposal. If only one proposal is funded, then a much reduced version of the whole undertaking will be possible. The ROLE proposal focuses on development of the micro-ethnographic methodology for HCI, and tries it out with three software systems: SimRocket, VIRTUALBIOLAB for college freshman, VIRTUALBIOLAB for distributed high school AP students. The ITR proposal focuses on the Perspectives technology for distance education and uses the micro-ethnographic methodology for assessment.
Project Schedule

The project focus is on the development of information technology for distributed collaborative learning. This centers on the design and development of WEBPERSPECTIVES, its integration with VIRTUALBIOLAB, and its eventual dissemination as Open Source software. During the school year, this software will be tested with small groups of college and high school students. Assessment of these trials will feed back into the software development. Project findings will be broadly disseminated. Below is a timeline for major phases of the project:

<table>
<thead>
<tr>
<th>Software Development</th>
<th>Micro-ethnographic Assessment</th>
<th>Dissemination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall '01</td>
<td>Development of WEBPERSPECTIVES</td>
<td>Pilot testing of VIRTUALBIOLAB</td>
</tr>
<tr>
<td>Spring '02</td>
<td>Extensions to WEBPERSPECTIVES</td>
<td>Testing of VIRTUALBIOLAB</td>
</tr>
<tr>
<td>Summer '02</td>
<td>Integration of WEBPERSPECTIVES with VIRTUALBIOLAB</td>
<td></td>
</tr>
<tr>
<td>Fall '02</td>
<td>Addition of negotiation</td>
<td>Testing of VIRTUALBIOLAB</td>
</tr>
<tr>
<td>Spring '03</td>
<td>WEBPERSPECTIVES maintenance</td>
<td>Pilot testing with WEBPERSPECTIVES</td>
</tr>
<tr>
<td>Summer '03</td>
<td>WEBPERSPECTIVES revision</td>
<td></td>
</tr>
<tr>
<td>Fall '03</td>
<td>Addition of hyper-linking</td>
<td>Testing with WEBPERSPECTIVES</td>
</tr>
<tr>
<td>Spring '04</td>
<td>System clean-up</td>
<td>Testing with WEBPERSPECTIVES</td>
</tr>
<tr>
<td>Summer '04</td>
<td>Open Source release</td>
<td>Report on analysis of testing</td>
</tr>
</tbody>
</table>

Plan for Assessment

We will engage in formative evaluation of our project throughout. That will be an important function of our Advisory Board, which includes assessment experts, and will form a regular part of the monthly workshops. We will check that we are making progress toward our project goals in accordance with the project timeline and are following our data analysis procedures.

The micro-analytic approach that the project will develop provides a built-in assessment process for the project. By videotaping sessions of students working with artifacts, we will derive a formative evaluation of the learning facilitated by the artifacts. By the end of the project, we will be able to compare in a detailed and documented way how well our revised versions of educational software artifacts perform as compared to how they worked in the pilot studies and in earlier phases of the project.

In addition to the micro-ethnographic analysis which examines both how students learn with computer technologies and their learning processes as revealed through their interactions (computer-mediated and face-to-face), it is important to understand how students relate to the technologies. In order to understand this, a triangulated approach to assessment will be adopted. Some students in the core trials of VIRTUALBIOLAB will be given a set of
pre-assignment questions to gauge their prior knowledge and understanding of the concepts. Once they have completed the trial, they will be asked the same questions so that we can calculate their learning gains. In addition, we will interview these students in order to understand their perceptions of the artifacts as effective learning tools. This information will be gathered with each iteration and use of the software under development, and the comments and perceptions will be fed back into the development of software and the articulation of learning processes that involve computer software and computer-mediated collaboration. Understanding student perceptions of their experiences will also enable us to track our progress toward our research goals and to evaluate the effectiveness of the software by answering the critical question of, does it work: have we effectively supported distance collaborative knowledge building?

Project Management

The PI will be personally responsible for coordinating activities associated with the project. He will supervise the work of students and consultants and ensure that they are working in accordance with the project plan, including the procedure for the collection and analysis of data. The PI will make certain that the plan is followed and the timetable met (taking into account changes adopted during the life of the project). He will also attempt to mediate any conflicts that arise within the diverse and interdisciplinary project staff. The PI will engage project Advisory Board consultants who are assessment specialists to assist in on-going project evaluation and to conduct a quarterly project review for reporting to the Advisory Board.

Data collection and analysis issues including sampling and confidentiality will conform to rigorous research conventions and University of Colorado Human Subjects standards.

Plan for Advancing Knowledge within Related Fields

We will establish a website for both internal use and broad dissemination. The website will collect and coordinate materials and findings of the project. It will include logs of our videotapes, digitized clips of selected episodes, detailed transcripts, analyses of interactions, etc. It will also include all papers submitted to journals and conferences.

This project and its findings will be broadly disseminated in the CSCL, CSCW, HCI, education, and communication research communities through conferences and journals. It will be particularly prominent at CSCL 2002 and subsequent meetings of CSCL, AERA, CSCW, Group, ICLS, and
WebNet. It will also significantly impact the release of a published VirtualBiologyLab curriculum at the college and the high school level.

The Project Advisory Board

The project team includes a number of Advisory Board members at the University of Colorado who bring important complementary skills and expertise from education, ethnography, human-computer interaction, and assessment. Science and math content expertise is provided by Kalmon, Nathan, and Otero. In addition, the developer of VIRTUALBIOLAB is involved (Klymkowski). Advisory board members attend monthly workshops to review project progress and to participate in data sessions. Individual members may take a more active role during specific project periods as needed. The breadth of the team is important to provide an interdisciplinary audience for the analysis of the data, as well as for the ongoing design and assessment (esp. Barker, Garvin-Doxas, Palen) of the project.

Lecia Barker is Director of Evaluation and Assessment for ATLAS, the Alliance for Technology, Learning and Science. Her PhD dissertation in Communication was on the discursive construction of virtual community in LAM DAM MOO, a computational artifact. She is currently Evaluator or co-PI in 7 grants, mostly from NSF.

Robert Craig is an Associate Professor of Communication who specializes in communication theory and discourse analysis. He has studied interactive discourse in college classrooms, especially critical thinking courses. He has written on theory of communication, grounded theory, and communication as a research field.

Kathy Garvin-Doxas conducts quantitative and qualitative assessment and evaluation of the SOLAR SYSTEM COLLABORATORY – a web-based freshman astronomy course for non-majors; for the past 3 years she has been evaluating the implementation of collaborative learning techniques at CU. Her PhD dissertation in Communication was a Bakhtinian analysis of collaboration in an organization becoming more participatory.

Stevan Kalmon is now a Regional Technology Consultant for the Colorado State Department of Education. Previously he was a high school teacher at New Vista H. S. in Boulder and educational advisor to L3D. He will be a liaison for working with high schools in the later part of the project.

Michael Klymkowsky is the designer and developer of the VIRTUALBIOLAB. He is a Professor of Molecular, Cellular &
Developmental Biology at CU. He has previously published textbooks and a CD-ROM (see his appended letter of support for details).

Curtis LeBaron is an Assistant Professor of Communication. He studied with Streeck and Hopper, founders of micro-ethnography, and has subsequently developed the field further and applied it to learning and to the analysis of artifact usage.

Mitchell Nathan is Assistant Professor in the CU School of Education. His specialty is learning and teaching in school settings, and educational technology. He conducts extensive field research in high school algebra classrooms. After earning his PhD in cognitive psychology, he did research at the Learning Research and Development Center at Pittsburgh and Vanderbilt University Learning Technology Center, where he worked on Jasper and Scientists in Action.

Valerie Otero joined the faculty of the CU School of Education in January 2001 as a science education specialist. She was a co-developer of curriculum materials and simulation visualization tools in a five-year NSF-funded project entitled: Constructing Physics Understanding in a Computer Supported Learning Environment. Her PhD dissertation was on the shifting role of visualization tools in the process of learning electrostatics in a collaborative environment.

Leysia Palen is an Assistant Research Professor in Computer Science and a member of L3D. She specializes in the adoption and use of everyday artifacts like mobile phones and electronic and group-sharing calendars. She studied CSCW issues and ethnographic methods at UCSD and UC Irvine.

Tamara Sumner is Assistant Professor of Computer Science at CU and is a member of L3D and ICS. She teaches HCI, AI, and the Internet. She is co-founder and co-editor of JIME. Her research includes digital libraries in geo-science and on-line scholarly publication. Previously she developed distance education courses at the Open University in England.

**Investigation of Innovative Concepts**

The project centers on exploring the use of computational Perspectives to support collaboration in distance education. The project’s theory argues that collaboration is essentially structured by the interaction of personal and group perspectives. A system that can mirror or represent this structure may be able to provide valuable support for the difficult process of collaboration in learning. The theory of knowledge building also argues that collaboration is crucial to the kind of learning that needs to take place in science education. Although there seems to be a tremendous potential for advances
in the quality of distance education through Perspectival support of collaboration, the effort faces serious hurdles. In particular, our pilot studies show that adoption and motivation are difficult issues that confront the design of this software. The project will therefore adopt a micro-analytic approach to studying how small groups of students interact with different versions of WEBPERSPECTIVES and VIRTUALBIOLAB.

Project Context and Resources

The proposed project will have considerable institutional support from the University of Colorado. The PI is a Research Professor active in both the Institute of Cognitive Science (ICS) and the Center for LifeLong Learning and Design (L3D). He is on the Steering Committee of the international SIGGroup conference, GROUP ’01, and is Program Chair of the international CSCL ’02 conference (both being held in Colorado). These affiliations will provide support and visibility to this project.

Through its Advisory Board, the project brings together a valuable set of experienced researchers from Education, Communication, Computer Science, and project assessment.

7. Expected Impact of Proposed Research

Expected Advances in Related Fields

The proposed research has implications for several fields. The WEBPERSPECTIVES software will be relevant to computer support for collaboration generally (CSCW and CSCL). The micro-ethnographic studies of VIRTUALBIOLAB with WEBPERSPECTIVES will provide insight into human-computer interaction, processes of collaborative learning, and methodology for assessing distance education software.

Expected Contribution to Teaching, Training and Learning

The project is focused on issues of teaching, training, and learning in distance education, particularly how to support collaboration in distance education.

The project will itself provide a teaching, learning, and training opportunity for the PI, Advisory Board members, graduate research assistants, and students in courses at the University of Colorado associated with the project through class projects.
Expected Broadening of Participation

Distance education offers the possibility of participation in quality educational experiences regardless of gender, ethnicity, disability, or geographic region. Studies show that disadvantaged and socially oppressed people often feel freer to participate in computer-mediated interactions, where their personal characteristics are not as immediately visible. The real gain will happen when everyone can engage in serious collaboration without these traditional restrictions to participation. WEBPERSPECTIVES is designed to promote such participation in a way that individuality is not lost in the process.

Expected Enhancement of the Infrastructure for Research and Education

WEBPERSPECTIVES and VIRTUALBIOLAB are two direct contributions to the software infrastructure for education. The use of micro-ethnography provides a methodological contribution to research in education.

Expected Dissemination

Project results will be widely disseminated. WEBPERSPECTIVES will be made available as an Open Source component; the PI has already had requests from California and Germany for this. VIRTUALBIOLAB will be published by a commercial textbook publisher in both college and high school versions. This project will significantly improve the quality of this distance education curriculum by subjecting it to extensive and detailed assessment in use situations. The addition of WEBPERSPECTIVES to this curriculum will add an important collaborative dimension to it.

Expected Benefits to Society

Science education for the general public, such as biology for non-specialists, is of obvious importance in our technological age, especially as genetic technology starts to change daily life. As we become an increasingly global society, the ability to collaborate freely over computer networks assumes increasing urgency. We are still far from understanding how to support collaboration. This project explores one promising mechanism: the representation of knowledge-building perspectives.
References


Proposals for Research


Streeck, J. (1983) Social Order in Child Communication: A Study in Microethnography, Benjamins, Amsterdam, NL.
The Role of Computational Cognitive Artifacts in Collaborative Learning and Education

This project addresses ROLE quadrants 2 and 3: It builds bridges from cognitive and social theories of the role of artifacts to research on learning in educational settings, and it develops a methodology for the principled assessment and research-based design of technological artifacts to mediate learning processes. The goal of the project is to refine both a micro-analytic methodology and an artifact-centered theoretical framework that can aid in the principled design of distance learning environments. The project will not only result in a much-needed methodology for future designers of educational technology, it will also deepen our understanding of the role that such computational cognitive artifacts can play in collaborative learning and formal education.

The project studies small groups of students using prototype versions of learning environments to see what the students go through in learning how to use the computer-based artifacts and what problems interfere with the learning goals. The project brings together educational software developers and experts in human-human and human-computer interaction to conduct the analysis of videotaped student interactions and to iterate the design of the educational environments.

Three software systems developed by project team members are studied, gradually advancing from a relatively simple computer simulation, through a semester-long on-line biology lab curriculum, to a distance education version of the labs:

1. SIMROCKET simulates the launch of rockets having varying characteristics. Five middle school students used the simulation to predict the effects of the different characteristics on the height attained by the rocket. Their sessions working with the simulation were videotaped.

2. VIRTUALBIOLOGYLAB is a series of 10 freshman college biology lab experiments simulated and conducted on the Web. Groups of 2 or 3
students work together to complete the lab – analysis of their successes and difficulties feeds back into the iterative design of the digital curriculum.

3. WEBGUIDE is a knowledge-building environment for supporting collaboration. Functionality from WEBGUIDE will be integrated into the VIRTUALBIOLOGYLAB in the final project year, and used by geographically distributed high school students for a version of the biology labs redesigned for Advanced Placement study.

From the perspective of the project’s theoretical framework, these learning environments are treated as interacting networks of computational cognitive artifacts. For instance, in analyzing students working with SIMROCKET, the project team looks at how the students talk about and make use of (a) the rocket simulation, (b) a display of rocket characteristics, and (c) a data collection form. Each of these three artifacts is designed in a way that permits it to be used in certain ways to accomplish certain tasks: (a) some artifacts like the simulation are computational and change on their own in response to inputs; (b) others like the display convey knowledge; while (c) yet others like the form provide cognitive support by organizing and preserving information. In each case, the students must learn how to recognize and take advantage of these artifact affordances. Designers of learning environments must design both the affordances of individual artifacts and the curricular context that will make these meaningful to students within a coherent educational experience.

The project analyzes the collaborative efforts of small groups of students in order to determine (a) the extent to which students can understand the use of educational artifacts, (b) where this is problematic, and (c) how student learning can be scaffolded to overcome problems. In collaborative interactions, students must display to each other their beliefs, their questions, their problems, and the resolution of problems. When this process is videotaped and carefully analyzed, it makes the students’ learning visible to researchers as well. The project adopts a micro-analytic form of communication analysis called micro-ethnography to study what is displayed. This is a rigorous method for analyzing both vocal and visible forms of human interaction recorded on video.

The project team has already begun to adapt micro-ethnography and the theory of artifacts to the analysis of student interactions with on-line educational technologies. Members of the project team have collaborated in various combinations in the past, including a semester-long pilot project investigating SIMROCKET student interaction data and theories of artifacts. The 4 Principal Investigators are experienced in the design of educational
technology and/or the micro-ethnographic analysis of people interacting with educational artifacts. All 9 of the Advisory Board consultants assessed educational technologies in their PhD dissertations and/or in their current work. The 2 Graduate Research Assistants are pursuing dissertations closely linked to this project.

1. The Problem of Educational Artifacts

As schools across the nation get wired for computer-based learning, the problem of scarcity of effective on-line curriculum and content becomes increasingly urgent. Current research stresses the importance of carefully designed software artifacts that are “student centered, knowledge centered, assessment centered, and community centered” (Bransford et al., 1999). Yet, with a few notable exceptions, there is little in the way of constructivist curriculum and content that meets these criteria and is also ready to take advantage of computer-and Internet-based media. In fact, there is little systematic knowledge of how to develop such educational artifacts, grounded in a theoretical understanding of the role that such artifacts might play in learning and in a methodology for software testing.

Distance Learning

The general problems of computer-mediated education multiply substantially under the pressure to rush to distance learning. Universities and dot.com’s around the country have jumped on the distance education bandwagon, without necessarily thinking through the complex educational issues involved. It is true that distance education has the potential to address various pressing educational, social, geographic, and economic issues (Keegan, 1986). It is also true that the technical infrastructure that will enable this revolution in education to proceed is being quickly set in place. However, the design and development of the necessary curriculum and content lags far behind. In the dash to market, providers of distance education are likely to settle on software technologies that were developed for other uses and are inappropriate for educational applications, curricula that implement outmoded approaches like drill-and-practice, and content that has not been tested for its learning effects. We need to develop more new models of computer support for learning that are effective in distance learning.
Designing Computer Support for Learning

Based on our own experiences with software in classrooms, we have found that computer-supported collaborative learning (CSCL) has a vast – and largely untapped – potential. Access to global sources of information is just one facet. In addition, computer simulations can transform conceptual representations into interactive worlds for inquiry. They can transcend real-world barriers of time, expense, geography, scale, expertise, etc. to allow students to engage with and experience phenomena that have until now been unapproachable. Hypertext systems of information can personalize presentations to meet individual learning needs. Communication media can promote collaboration in ways never before possible, as well as among people who could not hitherto interact. Structured curricular databases and shared knowledge-building environments can support student learning processes. However, we have seen that students always use computer artifacts in ways not envisioned by the designers. So, careful study of the artifacts in naturalistic settings is critical to the development of effective educational technology.

Understanding Computational Cognitive Artifacts

It is possible to ground the design and assessment of educational software applications in an understanding of their role as “artifacts” in learning. Our preliminary understanding views various forms of artifacts as absolutely central to human cognition and learning. People construct their understanding through interaction with artifacts; often artifacts extend, amplify, or transform cognition; eventually the artifacts may be internalized as mental procedures (Cole & Griffin, 1980; Donald, 1991; Engelbart, 1995; Hutchins, 1999; Norman, 1993; Papert, 1980; Pea, 1985; Vygotsky, 1930/1978). We intend to further develop this theoretical framework, which is inherent in theories of situated action and situated learning, in distributed cognition, in activity theory, and in various philosophies.

In particular, we propose to apply this framework to analyze educational technologies as “computational cognitive artifacts” (CCAs). We use this term to refer to computer-based or Internet-based educational artifacts (simulations, data analysis tools, on-line curricular modules, etc.): They are “computational” if they respond interactively to user interactions by changing their display. They are “cognitive” to the extent that they can become part of the user’s thinking, by, for instance, helping the user to visualize some phenomenon, providing an external memory or workspace for manipulating representations, or aiding in conducting a computation. They may also be “cognitive” in the further sense that they can be
internalized in the user’s mind so that he or she can make use of them as a mental metaphor or representation in the future when they are no longer even virtually present on a monitor screen. They are “artifacts” in the sense that they are perceptible objects that were designed to serve as some kind of tool – even though today they may not be physical objects that can literally be grasped.

To conceptualize an educational application as a CCA is not to assume a priori that it functions effectively in this role. Rather, it is to raise a set of critical issues:

- Does it facilitate human-computer interactions or mediate human-human interactions computationally?
- Does it support and enhance cognitive functions of its users?
- Does it function as a useful artifact in accordance with its design?

By conceptualizing certain types of educational technology (e.g., computer simulations) as CCAs, we can ask if they are fulfilling this role effectively in specific situations that we observe. We propose to investigate how people develop the understanding required to use CCAs effectively in CSCL settings, and conversely to study the roles these CCAs then play within the collaborative learning and education taking place.

**Assessing Collaborative Learning for Iterative Design**

Our goal is to contribute to the design of CCAs that are effective for supporting collaborative learning and education. Our theoretical framework does not directly imply criteria for the design of educational technology. Rather, it suggests that we develop prototypes of software applications and look at how students actually relate to them as computational cognitive artifacts – that is, that we look at how students concretely explore, come to understand, and use the software as an artifact for extending their cognitive powers – and then we iteratively revise the design of the software. For us as researchers to look at this, we need a methodology. We believe that micro-ethnography provides such a methodology. Micro-ethnography was designed to look very closely at social interaction processes. This project will adapt micro-ethnography to look at computer-mediated interactions in situations of collaborative learning.

The proposed project is an application of micro-ethnography’s method to the concerns of human-computer interaction. It brings together a team of people from these areas who are experienced in interdisciplinary research (see Biographical Sketches). Our team includes faculty and students from
Communication, Computer Science, Cognitive Science, and Education, as well as developers of educational software – within a broader academic community that is supportive of this project. This project is unique in bringing together educational software developers and specialists in the micro-analysis of interaction to develop and systematically test a rigorous methodology and a grounded theoretical framework for the design of distance learning artifacts.

2. Three Educational Artifacts for Study

In our project we will study the use of three software systems that we have developed: SIMROCKET, VIRTUALBIOLOGYLAB, and WEBGUIDE (see section on Pilot Studies for more details):

A Middle School Computer Simulation

SIMROCKET is a simulation of rocket launches. We already conducted and videotaped a three-hour trial of it with 5 middle school students and a teacher. We have begun to analyze the data from this trial. We have observed that the simulation artifact played a central role in the interaction: it opened up and defined the whole educational space, providing the narrative context as well as the source of data for collection and interpretation. In the analysis of a specific episode with SIMROCKET (see below), we will see the collaborative interaction revolving around three inter-related artifacts: the computational simulation of rocket launches, an external memory display of rocket characteristics, and a paper chart of recorded rocket heights. By closely analyzing the interactions among the teacher and students we see: (a) successes and failures of students to grasp the meaning/use of these artifacts, and (b) the teacher’s attempts as an experienced scientist to guide the group to effective use of the artifacts. Because of problems in the interaction that become apparent to the participants, the teacher must make his analytic skills observable and the students must make their adoptions or misunderstandings apparent. We also see that there is not a single simple artifact here, but a subtle network of artifacts with different functions. Furthermore, the artifacts only exercise their cognitive function or activate their meaning when they are being used appropriately. Our observations of the teacher’s patterns of face-to-face interaction suggest forms of scaffolding that could be introduced in distance learning where a teacher is not physically present.
A College On-line Lab

The VIRTUALBIOLOGYLAB is a much more complex network of interdependent artifacts. It is a complete one-semester curriculum on the Web, intended to replace college freshman biology wet labs for non-majors. Each of 10 planned labs takes an estimated three hours for a student to work through – and enables students to conduct seminal experiments from the history of biology that would not be feasible in traditional physical wet labs (see attached letter of support from the developer). One can distinguish multiple kinds of artifacts composing the software: a guiding narrative, animations of lab equipment, simulations of lab procedures, data collection / analysis / graphing / display tools, background materials (theory, history, remedial text), links to related websites, and interactive assessment exercises. The virtual lab is designed to be used by students independent of any teacher guidance, although it is loosely coordinated with a biology lecture course. The on-line system must work as a whole, motivating and guiding students through a sequence of tasks; each of the distinct component artifacts must work effectively on its own and within the whole pedagogical context.

A High School Distributed Education Lab

WEBGUIDE is a knowledge-building environment to support collaborative learning. It provides a collaboratively constructed and shared external memory medium on the Web. The display is dynamically computed to show a hierarchy of notes arranged as a personal or group “perspective” on the persistent, asynchronous discussion. This perspective mechanism is an artifact that people must learn how to use and navigate to mirror and support the interpersonal relationships of collaboration. WEBGUIDE also incorporates a variety of knowledge management functions that must be learned in order to manipulate the ideas stored in the system and to build effective shared knowledge. Certain components of WEBGUIDE will be integrated with a version of the VIRTUALBIOLOGYLAB toward the end of our project to explore a collaborative distance learning biology curriculum at the high school Advanced Placement level. We will also extend the lab software to incorporate educational scaffolding techniques from other knowledge-building environments like CSILE/KNOWLEDGEFORUM (Scardamalia & Bereiter, 1996), KIE/WISE (Cuthbert, 1999), and COVIS (Pea, 1993).
3. A Staged Research Plan

Collaborative learning is a complex process. Accordingly, our project will build up gradually from our relatively simple pilot study to a full example of collaborative distance education.

Project Schedule

The project will consist of three main stages:

1. Analysis of the three hours of video tape already collected of five middle school students and a teacher conducting a study of rocket design using the SIMROCKET computer simulation.

2.a. A very brief study of college freshmen in a biology wet lab. This will serve as an informal baseline for the next stage.

2.b. This is the core study for the project. We will videotape small groups of students working together with the on-line VIRTUALBIOLOGYLAB. This software is currently under development at the University of Colorado. The developers are involved in this project and will be iterating the design of the software in response to the analysis of the usage trials. We will focus our analysis on critical steps in the lab, like learning how to use a particular piece of equipment or a data analysis tool.

3. A distance education version of VIRTUALBIOLOGYLAB will incorporate a collaboration medium based on WEBGUIDE. This will be offered as an Advanced Placement curriculum to geographically distributed high schools students. The curriculum will be designed to be collaborative, and we will log user interactions and use these to study the learning taking place.

Following is a timeline for these stages:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Summer '01</td>
<td>data analysis</td>
<td>2.a. &amp; 2.b. pilot trials</td>
<td></td>
</tr>
<tr>
<td>Fall '01</td>
<td>complete data analysis</td>
<td>collect data</td>
<td></td>
</tr>
<tr>
<td>Spring '02</td>
<td>revise method</td>
<td>iterate &amp; collect data</td>
<td></td>
</tr>
<tr>
<td>Summer '02</td>
<td>revise theory</td>
<td>data analysis</td>
<td>pilot trials</td>
</tr>
<tr>
<td>Fall '02</td>
<td>iterate, collect, analyze data</td>
<td>collect data</td>
<td></td>
</tr>
<tr>
<td>Spring '03</td>
<td>iterate, collect, analyze data</td>
<td>collect data</td>
<td></td>
</tr>
<tr>
<td>Summer '03</td>
<td>complete data analysis</td>
<td>complete data analysis</td>
<td></td>
</tr>
<tr>
<td>Fall '03</td>
<td>revise method &amp; theory</td>
<td>revise method &amp; theory</td>
<td></td>
</tr>
<tr>
<td>Spring '04</td>
<td>evaluate project</td>
<td>disseminate findings</td>
<td>prepare final report</td>
</tr>
</tbody>
</table>
Data Gathering and Analysis

Our gathering and analysis of data involves the PIs working closely with the graduate and undergraduate team members. In addition, our consultants participate in workshops held monthly. The workshops not only review project progress and plan next steps, but they importantly include group data sessions for the analysis of data. The data gathering and analysis process (for instance for the VIRTUALBIOLOGYLAB sessions) will typically proceed through the following steps:

0. Videotaping of students. Two or three students are gathered around a computer. Cameras and microphones are set up to capture the facial expressions and body movements of all participants. The monitor image is also captured. Microphones are arranged to capture all speech as clearly as possible and to distinguish the speakers.

1. The video is combined (picture-in-picture) and time-code is burned in to provide a frame-by-frame reference system.

2. A minute-by-minute record log is created, describing in a sentence or two what takes place each minute. This is typically done by a graduate student and reviewed by a PI. The log may be revised later.

3. A list of interesting episodes is created. Episodes are meaningful interactions lasting up to several minutes. The list is discussed by the whole project team at a group workshop.

4. Selected episodes are digitized and made available electronically. This allows them to be replayed easily, looped, freeze-framed, slowed down, and studied by project consultants at distant locations.

5. A detailed transcript is created. It transcribes both speech and visible behaviors. Speech of different participants is color-coded. The transcripts are printed and posted on the Web with the digitized clips.

6. Each episode is assigned to a project team member who “owns” that piece of data. The owner watches the clip many times to understand what is happening there.

7. A data session is conducted with the whole project team at a group workshop. This is a collaborative analysis of the data’s empirical details. Usually, about two hours are spent on a single episode. The session is led by the owner of the data, who presents the episode and raises issues. The owner may audio-tape this session to preserve ideas and interpretations that come up.
8. The owner of the episode returns to a study of the video clip. At this point, the transcript may be revised and extended to include more details of interaction. The owner may invite other project team members to view and discuss the clip. The owner may present the clip at another data session. Finally, the owner drafts a micro-ethnographic analysis of the episode. This is distributed for comment. The analysis includes:

   a. A detailed description of the actions of all participants and their interactions.
   b. A discussion of what learning is evidenced in the data.
   c. A discussion of the role of any artifacts.
   d. A discussion of problems with the software, learning problems, etc.

0. The analyses of the episodes are reviewed by the whole project team and various suggestions are made based on this:

   a. Proposed revisions to the software.
   b. Changes to the list of interesting episodes, such as the inclusion of additional episodes.
   c. Alterations to the research plan, such as scheduling additional usage sessions or changing the way they are conducted.
   d. Revisions to the research methodology and theoretical framework.

**Project Assessment and Dissemination**

We will engage in formative evaluation of our project throughout. That will be an important function of our larger team, which includes assessment experts, and will form a regular part of the monthly workshops. We will check that we are making progress toward our project goals in accordance with the project timeline and are following our data analysis procedures. Specifically, we will check that we are developing our methodology for making learning visible and for iteratively designing software artifacts, as well as disseminating our findings.

The micro-analytic approach that the project will develop provides a built-in assessment process for the project. By videotaping sessions of students working with artifacts, we will derive a formative evaluation of the learning facilitated by the artifacts. By the end of the project, we will be able to compare in a detailed and documented way how well our revised versions of
educational software artifacts perform as compared to how they worked in the pilot studies and in earlier phases of the project. In addition, we will assess how successful we were in the course of the project in developing, formulating, and applying micro-ethnographic methodology for studying the educational role of cognitive artifacts and for assessing the ability of students to adopt the computational artifacts into their collaborative learning.

In addition to the micro-ethnographic analysis which examines both how students learn with computer technologies and their learning processes as revealed through their interactions (computer-mediated and face-to-face), it is important to understand how students relate to the technologies, as well as the degree to which students learn. In order to understand this, a triangulated approach to assessment will be adopted. Some students in the core trials of VIRTUALBIOLOGYLAB will be given a set of pre-assignment questions to gauge their prior knowledge and understanding of the concepts. Once they have completed the trial, they will be asked the same questions so that we can calculate their learning gains. In addition, we will interview these students in order to understand their perceptions of the artifacts as effective learning tools. This information will be gathered with each iteration and use of the software under development, and the comments and perceptions will be fed back into the development of software and the articulation of learning processes that involve computer software and computer-mediated collaboration. Understanding student perceptions of their experiences will also enable us to track our progress toward our research goals and to evaluate the effectiveness of the theory and method under development by answering the critical question of, does it work: have we indeed made learning visible in a way that can contribute to iterative design of effective software artifacts?

The PI will be personally responsible for coordinating activities associated with the project. He will supervise the work of students and consultants and ensure that they are working in accordance with the project plan, including the preceding procedure for the collection and analysis of data. The PI will make certain that the plan is followed and the timetable met (taking into account changes adopted during the life of the project). He will also attempt to mediate any conflicts that arise within the diverse and interdisciplinary project staff. The PI will engage project Advisory Board consultants who are assessment specialists to assist in on-going project evaluation and to conduct a quarterly project review for reporting to the Advisory Board.
Data collection and analysis issues including sampling and confidentiality will conform to rigorous research conventions and University of Colorado Human Subjects standards.

We will establish a website for both internal use and broad dissemination. The website will collect and coordinate materials and findings of the project. It will include logs of our videotapes, digitized clips of selected episodes, detailed transcripts, analyses of interactions, etc. It will also include all papers submitted to journals and conferences.

This project and its findings will be broadly disseminated in the CSCL, CSCW, HCI, education, and communication research communities through conferences and journals. It will be particularly prominent at CSCL 2002 and subsequent meetings of CSCL, AERA, CSCW, Group, ICLS, and WebNet. It will also significantly impact the release of a published VIRTUALBIOLOGYLAB curriculum at the college and the high school level.

4. Theoretical Framework for Cognitive Artifacts

In the current Fall 2000 semester, the PI offered an interdisciplinary seminar on the theory of artifacts. Many of the project co-PIs, graduate research assistants, and consultants participated fully in the seminar. The project’s theoretical framework grow out of this seminar. It will be considerably refined through a grounded theory analysis of the data collected in the project. This theoretical framework provides a bridge from selected findings of various cognitive sciences to research on learning in educational settings. It will guide the questions we pose in looking at our data.

Mediated Cognition

We start from three principles enunciated by Vygotsky (1930/1978; 1934/1986):

1. Mediated cognition. Modern human cognition is thoroughly mediated by physical and symbolic artifacts such as tools and words. We extend this to the use of computer-based artifacts like simulations, data analysis tools, and collaboration media.

2. Social cognition. Meanings and practices are first established interpersonally and may then be internalized in individual minds. We take advantage of this by analyzing the interpersonal interactions, which
are largely observable to the trained analyst as well as to the participants.

3. **Zone of proximal development.** A student learns most productively when guided somewhat beyond his or her current developmental level by peers or a mentor. We use this principle to design experimental situations in which a small group of students is challenged to engage in a scaffolded scientific task.

**Collaborative Knowledge Building**

We conceptualize our subject matter as the process of “knowledge building” (Bereiter, 2000). This is an active collaborative learning process in which a community constructs conceptual meaning. For instance, in our SIMROCKET pilot study the students came to understand the effect of different variables upon future rocket launches and learned to isolate variables to measure their independent effects. The process of collaborative knowledge building is interpersonal and observable – primarily through analysis of the communicative interactions through which it takes place.

Collaborative knowledge building involves an interplay between individuals and the group, with individuals contributing from their personal perspectives and the group accepting these contributions in its own way (Stahl & Herrmann, 1999). This perspective-taking and perspective-making unfolds in the observable world of signs and artifacts, such as spoken utterances and external memory devices (Boland & Tenkasi, 1995). The physical and symbolic artifacts mediate between personal and group understandings.

**The Role of Artifacts**

It is possible to re-conceptualize learning (both individual and collaborative) through a focus on the artifacts that are involved. Artifacts – including software artifacts – embody intentionality, meaning, and experiences of their creators and preserve these for future users (Donald, 1991; Hall, 1996). The problem is for users of artifacts to know how to re activate this stored wisdom. This requires complex skills of interpretation (Gadamer, 1960/1988; Stahl, 1993). Education can be viewed as largely the effort to socialize children and other new-comers into a practical understanding of the artifacts and practices that constitute a society’s or a community’s culture (Lave & Wenger, 1991). The written word and the symbols of mathematics, for instance, are cognitive artifacts that take years of schooling to master. While people have been producing and using artifacts forever
Artifacts play an absolutely central role in learning and understanding according to the philosophic roots that underlie contemporary cognitive theories that are influential for CSCL theories (Koschmann, 1996; Koschmann, 1999; Koschmann, in press), such as situated action (Suchman, 1987), situated learning (Lave & Wenger, 1991), activity theory (Engeström et al., 1999), distributed cognition (Hutchins, 1996), dialogicality (Bakhtin, 1986), and critical inquiry (Dewey & Bentley, 1949/1991).

According to Hegel (1807/1967), the very basis of self-consciousness and sociality in mutual recognition is thoroughly mediated by the creation and use of artifacts – which embody human consciousness or meaning in their imposed form or design. Marx (1867/1976) argues that the production, circulation, and consumption of artifacts as commodities is both affected by the prevailing social relations and reproduces those relations – and influences how we understand and learn about contemporary artifacts; these commodities are essentially stored labor – physical and intellectual – that comes alive in use. Marx traces the social history of artifacts from simple tools through machinery to computational automated industry. For Husserl (1936/1989), meaning is established and historically sedimented in the form of artifacts; Heidegger (1927/1996) expands this analysis to argue that the life-world of our everyday involvements is structured as networks of meaningful artifacts. More recently, software is seen as a new form of stored meaning or intentionality (Keil-Slawik, 1992; Stahl, 1993; Winograd & Flores, 1986). For instance, effects of “artificial intelligence” are accomplished by embedding human intelligence in software procedures and knowledge-bases.

Engelbart (1995) and Norman (1993) claim that it is artifacts that make us smart, by amplifying our very limited native abilities like short-term memory. Others (e.g., Cole & Griffin, 1980; Pea, 1985) counter that these artifacts change our tasks, rather than simply increasing our powers, but this still places artifacts centrally in our attempts to increase our intellectual capabilities. Donald (1991) argues that the entire enterprise of modern knowing and science only became possible with the development of artifacts like books, which provided external memories that could be circulated and that might outlive their creators. Papert (1980), reflecting on his own learning history, believes that playing with automobile gears as a young child “did more for my mathematical development than anything I was taught in elementary school. Gears, serving as models, carried many otherwise abstract ideas into my head” (p. vi).
If one looks closely at learning – from infancy to kindergarten, formal schooling, and on-the-job – one sees that artifacts (now including computational artifacts) are pervasive. While it is clear that a primary function of education (and socialization into culture generally) is to teach new-comers how to understand and use the available artifacts of one’s society or of its specialties, we have only narrow studies of how this takes place. For instance, Bruner (1990) discusses how children acquire the ability to follow and generate narratives as verbal cognitive artifacts, and Hall (Hall & Stevens, 1995) investigates how young students use design tools.

**How Artifacts are Understood**

Even in our very preliminary pilot study of the SIMROCKET data, it has already become clear that the process of coming to understand a computer simulation that models a scientific phenomenon is a complex process, which strains the cognitive abilities of middle school students. Without strong guidance from a teacher, the students would at best have treated the simulation as a video game, perhaps competing to get the highest rocket flight, but not investigating the scientific factors that might lead to success.

Although students often make statements that sound like they understand how to construct certain kinds of knowledge, when one watches them struggling through the steps that are actually required one gains a much more detailed understanding of what is involved for a novice, what supports are helpful, and where problems typically arise. For instance, while the students in the pilot study were proficient at taking averages of sets of numbers in a traditional math lesson, they ran into many problems when averaging their rocket data. A major problem had to do with the organization of the data and of their averages on a data sheet. The two teams of students became very confused about which rocket heights had been observed by which team, and which averages were associated with them. While an adult experienced with scientific experiments can keep these things straight without thinking about it, the students had to learn this skill. They did this partially by negotiating with the teacher, who alerted them to problems and guided them back on track, and partially by collaboratively applying their own intellectual and communicative skills.

Our work and that of our current and past colleagues explores the use of gesture in understanding artifacts and in constructing shared understanding of artifacts. In his seminal example of micro-ethnographic analysis (which studies the interaction of five young children in a school project, and thereby provides a model for us) and subsequently, Streeck (1983; 1993; 1996) focuses on the roles of gesture in making social understanding visible.
LeBaron analyzes different forms of gesture that are successively used to build a shared vocabulary of meaningful gestural artifacts (LeBaron, 1998; LeBaron & Hopper, 1997; LeBaron & Koschmann, 1999; LeBaron & Koschmann, 2001; LeBaron & Streeck, 2000). Koschmann also highlights the role of gesture in educational settings (Koschmann et al., 1997; Koschmann & LeBaron, submitted; Koschmann et al., 1998; Koschmann & Stahl, 1998). Our micro-ethnographic method (see below) is explicitly adapted to making learning visible by systematically attending to the sorts of gestures and bodily interactions that people use to co-construct the meaning of artifacts.

According to our theoretical framework, learning through interaction with artifacts is an inherently social process, involving either interaction with other people through the artifact or at least interacting with an artifact that was made by other people and that incorporates their intentions. For our research, collaborative interactions have an important characteristic: in order to collaborate, participants must make their ideas and their relationships visible to each other as part of their communication. That is, they make learning visible. As researchers, we can capture this in video or computer logs and analyze it. That way, we can see how students are relating to computational artifacts and what they are learning in the process. This overcomes the traditional problem of educational assessment, where it is assumed that learning is invisible to researchers and must be inferred from learning outcome measures. Thus, our approach avoids the restriction of educational assessment to the kinds of analyses of pre/post-test statistics and after-the-fact interviews that so often lead to “no significant difference” (Russell, 1999) results, which are of little value for design purposes.

Of course, not all learning is made visible, so other methods to indirectly measure learning outcomes are necessary and complementary. But focusing on the visible displays of learning prevents the common tendency to lose track of the learning in favor of secondary phenomena that seem easier to describe or quantify. For instance, much of the traditional literature on cooperative learning focuses on small group facilitation, rather than on cognitive and group learning processes (for a recent review of this literature, see (Brody & Davidson, 1998) reviewed by the PI (Stahl, 2000a)). Even recent CSCL studies often miss the interesting learning phenomena (e.g., (Hakkarainen & Lipponen, in prep) and (Jong et al., in prep), reviewed by the PI (Stahl, in prep)).
Grounded Practical Theory

While we have encountered many suggestive ways of thinking about artifacts in our readings, the roles and functioning of artifacts are most clearly revealed by close observation of our data. We expect to come to a deep understanding of the role of artifacts in education – and conversely of the role of learning in artifacts – through our study of student interactions with educational artifacts.

Glaser & Strauss (1967) have described techniques for deriving theory from qualitative data in sociology. In philosophy, Gadamer (1960/1988) has proposed that hermeneutic understanding can be derived through reflection on life experience and situated interpretation. Schön (Schön, 1983; Schön, 1987) argues for reflective practice in professional activities like design and teaching.

Project co-PI Craig and his colleagues (Craig & Sanusi, 2000; Craig & Tracy, 1995) – building on Glaser & Strauss, Gadamer, Schön, and others – have developed an approach to grounded practical theory within communication analysis. The general idea of this approach is that practical theory – theory designed to inform praxis – involves conceptually reconstructing practice. This can be done on three levels:

1. A problem level that accounts for difficulties or dilemmas typically encountered in the practice.
2. A technical level that describes a repertoire of practical techniques for addressing problems.
3. A philosophic level that formulates normative principles to govern the use of techniques.

For example, collaborative learning is a normative principle that can govern the use of practical techniques such as the SIMROCKET exercise. A problem noted in the pilot project was that middle school students may not collaborate toward certain desirable learning objectives without some guidance by the teacher (level 1). In the SIMROCKET data, we see a teacher using various interactional techniques that may display his orientation to this problem. To facilitate reflection on those techniques, the problem might be conceptualized theoretically as an instance of the more general dilemma of any pedagogical practice that attempts to be learner-centered while achieving specific learning objectives. "Scaffolding" names a general sort of technique that teachers can use to address this dilemma (level 2), but scaffolding can be, for example, either too directive (becoming teacher centered) or too nondirective (risking failure to achieve prescribed
The collaborative learning principle (level 3) suggests a solution to the dilemma: the use of scaffolding techniques that focus the group's attention on a task that both structurally entails the prescribed learning objectives and requires active student collaboration to be completed. This may provide a principled basis for assessing the teacher's techniques in the SimRocket data, and also a principled basis for design revisions in the computational artifact (to better enable preferred forms of scaffolding). By the same token, the micro-ethnographic analysis provides a basis for assessing the relevance and applicability of this or any other theoretical reconstruction of the practice that might be proposed.

Such a grounded practical theory approach will guide us to:

1. Reflect upon problems that arise in the interactions we observe.
2. Define techniques that are responsive to these problems.
3. Formulate principled ways to move from empirical observations to software recommendations.

5. Research Methodology for Studying Interaction

Iterative Software Design

The core of the project is to develop a methodology for driving the iterative development of software for computer-mediated education. The idea is to start with an initial prototype, videotape small groups of students collaborating with the software, analyze the problems that arise as well as the kinds of learning that take place, formulate revisions to the software based on that analysis, and iterate system design (along with any associated recommendations for classroom presentation) toward improved learning.

Iterative design is a well-established approach in software development, particularly when the effectiveness of the software depends upon the ability of people to use it as intended. The problem is how to analyze the quality of usage in successive trials. This is best done by interpreting in a rigorous way how learning is taking place. During the past 25 years, scientific methodologies for interpreting social interaction have been developed. We focus on one particularly promising school of this science, micro-ethnography.
Micro-ethnography

For this project, we adopt a recent tradition of human interaction analysis (Jordan & Henderson, 1995) that we refer to as “micro-ethnography.” This methodology builds on a convergence of conversation analysis (Sacks, 1992), ethnomethodology (Garfinkel, 1967), nonverbal communication (Birdwhistell, 1970), and context analysis (Kendon, 1990). An integration of these methods has only recently become feasible with the availability of videotaping and digitization that records human interactions and facilitates their detailed analysis. It involves close attention to the role that various micro-behaviors – such as turn-taking, participation structures, gaze, posture, gestures, and manipulation of artifacts – play in the tacit organization of interpersonal interactions. Utterances made in interaction are analyzed as to how they shape and are shaped by the mutually intelligible encounter as a holistic context – rather than being taken as expressions of individuals’ psychological intentions or of external social rules (Streeck, 1983). At the same time, micro-ethnography addresses larger social concerns, such as criminal justice (LeBaron & Hopper, 1997), medical education (LeBaron & Koschmann, 1999), and problem solving in complex technological settings (Hutchins & Palen, 1998).

Micro-ethnographic research typically involves the following components:

1. A specific setting, or research site – such as several students gathered around a computer running specific software.

2. A detailed analysis of both audible and visible micro-behaviors, which are to be understood in terms of their embeddedness within the particular social and material environment – such as a classroom.

3. A recognition that culture (which includes the meaning and use of shared artifacts) is a product and a process of naturally-occurring communication, simultaneously co-constructed and experienced by participants – and thereby made available for empirical study and interpretation by researchers.

4. A use of recent technologies, like digitized video, that allow researchers to look at in detail the orderly performance of social life – such as the negotiation of learning between teacher and student or among collaborating peers.

We will build on this micro-ethnographic approach and on the expertise and methodology which has evolved through the micro-ethnographic data sessions conducted by the project co-PIs and their colleagues for several
years. We will collect appropriate data and conduct our own data sessions for project staff, as we have already begun to do with our pilot study data.

Micro-ethnography can be adapted from the study of human-human interaction to that of human-computer interaction or computer-mediated collaboration. Our pilot studies suggest that such an adaptation of the methodology can be accomplished effectively. Our past use of micro-ethnography in collaborative educational settings – particularly in medical problem-based learning – has been very insightful and encouraging.

**Micro-ethnography and Human-Computer Interaction**

Our research approach brings together educational software designers and micro-analytic researchers. We use micro-ethnography to analyze empirical student interactions with educational software artifacts. Techniques related to micro-ethnography, such as video analysis and conversation analysis, have previously been used to analyze human-computer interaction in limited cases (Bødker, 1989; Bødker, 1996; Frohlich & Luff, 1990; Hollan et al., 2000; McIlvenny, 1990; Nardi, 1996; Suchman, 1987; Suchman & Trigg, 1991). However, these cases typically did not analyze interactions at the micro-behavior level, including such things as gesture and posturing, which are important means of making understandings visible in face-to-face communication (suggestive exceptions from our own community include (Hutchins & Palen, 1998; Streeck, 1996)). But, most importantly, these studies did not investigate learning technologies. Nor did they investigate learning taking place through the interactions. Those that did look at learning (like Roschelle (1996)) did not use this to feed back into the design of the technology. Thus, our project is undertaking an approach that is unique in combining all three:

1. Analysis of interaction at a micro level.
2. Analysis of the learning taking place.
3. Application of the analysis to revision of the technology.

Our own past work using micro-ethnography has begun to move this approach toward our project goal. Co-PI LeBaron (1998) shows through micro-ethnography how an architecture teacher goes through four stages of successive abstraction to define meaningful gestures, which the students then gradually adopt in their own presentations. By freezing key video frames and relating them to the speech and bodily behaviors of the teacher and students, LeBaron makes the teaching and learning process – which the participants are only tacitly aware of – visible to researchers.
Co-PI Koschmann and collaborators (including LeBaron) have been engaged for almost 10 years in fine-grained studies of collaboration among medical students in a problem-based learning (PBL) curriculum (Glenn et al., 1999; Koschmann & Glenn, submitted; Koschmann et al., 1997; Koschmann et al., 2000; LeBaron & Koschmann, 2001). In particular, we have shown how group discussions raise learning issues for further study and how the status of these issues is negotiated by the students and a tutor. While we have investigated the role of a tutor in face-to-face PBL sessions, we have only recently begun to study the role of computer-based artifacts and media in distance-PBL sessions (Koschmann & LeBaron, submitted; LeBaron & Koschmann, 2001). The proposed project will build upon the isolated pioneering efforts of ourselves and others, and attempt to put these methods together in a systematic way and apply them to the design of educational artifacts.

Focus on Artifacts in Learning

Artifacts frequently play a central role in learning situations as analyzed by micro-ethnography. In the preceding examples, for instance, a student architectural model was the focus of discussion in LeBaron’s example, a medical instrument in Koschmann’s surgical example, and a whiteboard for listing learning issues in Koschmann’s PBL example. In each of these cases, the artifact was germane to the students’ learning. In fact, the students were primarily learning how to use and make sense of the artifact. The architecture teacher was demonstrating how to hold and study an architectural model, how to move through the spaces it creates, and how to critique its design. The surgeon was teaching his student how to manipulate the automated instrument and how to interpret the image on the screen as the organs being surgically treated. In the PBL session, the tutor was subtly guiding the students to formulate issues for the whiteboard and teaching them how to orient their collaborative learning processes around this artifact. In the following section, we see the role of artifacts in our SIMROCKET pilot study.

6. Sample Analysis of an Episode of Collaborative Learning

To illustrate our micro-ethnographic approach, we analyze a brief moment of classroom interaction taken from our SIMROCKET data. Our excerpt begins approximately 90 minutes into the videotaped record. The students have already performed multiple launches of seven different rockets, each
rocket having a certain combination of features (i.e., engine type, body type, nose shape, and number of fins). By noting the height of each launch, students were able to calculate an average height for each rocket. The students might have also compared these averages to determine the effects of the different features – but the students were having difficulty conducting such a comparison. Through his interaction with students, the teacher was able to assess participants’ understandings and lead them into a discussion of basic scientific procedures.

An Instance of Collaborative Learning with Artifacts

Consider the following transcribed moments, involving the teacher (T) and his various students (the transcript has been simplified for this presentation):

1   T:   And you don't have anything like that there?
2    (1.4) ((T gestures toward monitor & data))
3   S:   I don't think so.
4   J:   Not with the same engine. Not with the same•
5   T:   With the same engine but with a different nose
6    cone?

Repeatedly (e.g., lines 1 and 5), the teacher directs students’ attention toward their computer monitors (with the simulation and rocket description – see Figure 1 below) and their data sheets, inviting them to recognize these artifacts as having answers to the group’s various questions. One student (C) has just finished explaining how he would like to use computer software to drag various features onto simulated rockets for purposes of comparison. While looking and orienting toward C (line 1), the teacher gestures toward the computer monitors and data sheets (line 2), and thereby encourages the students to recognize their simulation artifacts as already embodying what they need for their questions. Having been prompted by the teacher, the students attend to what is the same or different about the simulated rockets in the description list. The transcription continues as follows:

7   C:   These are both of the same thing o•
8   B:   This one is different ((gestures toward monitor))
9   J:   Yeah but it has -uh
10  (0.4)
While the teacher remains silent, the students look toward their computer monitors and data sheets, and talk about the features of particular rockets. When C claims that two rockets are the same (line 7), B corrects by literally pointing out a difference (line 8). In collaboration with J, C notices that the rocket B referred to has a pointy nose cone (line 11) in contrast to rocket one’s rounded nose — still insisting, though, that “it’s not the same engine” (line 12). J and B quickly disagree with C (lines 13 and 14), and J prompts a comparison between rockets “two and one” (line 15). Through a series of discourse markers, C shows that he has a new understanding: the particle “oh” marks a change in his information state, and the recycled words “I see” provide additional evidence.
Through group interaction involving both vocal and visible forms of communication, participants' shifting understandings were made visible to each other (and to us), and through this process were eventually improved. C showed understanding of the need to compare rockets before he recognized the simulation software as the needed comparison activity. C's displayed vision of experimental design was essentially the same as the teacher's – C simply failed to see that the work of designing and data gathering had already been done. We are not claiming that this knowledge is something that individuals have acquired and firmly mastered; what interests us is the ability that they demonstrate in concrete interactional learning situations. Across videotaped episodes this knowledge is seen to be fragile and inconsistent, leading us to suspect that it may be dependent upon situational details, group dynamics, and knowledge-building processes.

Figure 1. The SimRocket interface consisting of the rocket simulation and the list of rocket characteristics.
Implications for Artifact Design

This analysis has implications for the design of the educational experience and of the artifacts that comprise it. The rocket simulation designer might have thought that putting the list of rocket characteristics so prominently on the computer screen would automatically enable the users to select rockets for comparison. The teacher actually began the session by having students read this list out loud; he might have thought that the reading and the carefully located list would make the implications unmistakable for the students. But we have just seen that it took a relatively extensive and collaborative effort at comparison before the structure of the information could become meaningful to the students such as C. They had to engage in particular ways with the simulation, rocket list, and data sheet artifacts within a collaborative task in order to learn something that a software designer might have taken for granted.

This interaction offers another insight with (re)design implications. C had offered an alternative to the simulation as he then saw it. In his effort to demonstrate his vision, he was actually very involved with the artifact on the screen, proposing dragging features to create rockets with the right set of features for comparing – even while not yet understanding the affordances that already existed in the given rockets. A designer who takes seriously the possibility that what C proposes might be more effective could consider making the design of the rockets an interactive aspect of the simulation. Rather than presenting rockets with already fixed sets of characteristics (which is what proved so hard for C to focus on), an interactive design would provide "parts" and require students to configure pairs of rockets with sets of characteristics that would allow them to measure the effect of different nose cones, etc. The results would be the same from a mathematical viewpoint, but the process might be more meaningful and insightful for the students and involve them interactively at an earlier stage of experimental design in scientific inquiry. It is also possible, of course, that it might be more confusing for students and chaotic for the group, but that is an empirical question to be decided after another trial and micro-ethnographic analysis.

7. Pilot Studies Conducted

The proposed project builds upon a series of activities that we have already started to work on. These activities – that either grew out of our previous engagements or were conducted to explore the basis for this proposal – have led to the design of our project and are suggestive of its probable success.

SimRocket (http://www.cs.colorado.edu/~gerry/previous/simrocket)

The PI designed and implemented SIMROCKET, a computer simulation of a rocket launch. The height attained by a simulated rocket is dependent upon its engine, nose-cone shape, fin configuration, and surface texture – as well as a random noise factor. The simulation was originally designed as part of a larger “Mission to Mars” curriculum built around launching model rockets and used with troubled middle school students in a remedial summer project. The PI was invited to try it in a local Boulder school with five boys engaged in a model rocket science project. The teacher guided the students (grouped in front of two computers) to fire each of 7 virtual rockets with different characteristics six times and to average the resultant heights in order to predict the height of an 8th rocket. Project staff then engaged in micro-ethnographic analysis of this three-hour interaction – during data sessions in the Communication Department, the seminar on artifact theory, a summer workshop on micro-ethnography, and pilot sessions for this proposal.

WebGuide (http://www.cs.colorado.edu/~gerry/webguide)

WEBGUIDE is a knowledge-building environment for discussing topics via the Web developed by the PI and colleagues over the past two years. It has been used in a middle school environmental science class and in college seminars on CSCL (Stahl, 1999c). WEBGUIDE goes beyond similar discussion-based systems by supporting the representation and development of personal and group perspectives (Stahl, 1999a; Stahl, 1999b).

VirtualBiologyLab (http://www.virtuallaboratory.net)

VIRTUALBIOLOGYLAB is a Web-based curriculum to substitute for biology wet labs in introductory freshman biology courses for non-majors. Currently under development, it will soon consist of ten labs, each of which takes approximately three hours for a student to work through. Project staff has begun to review this software with the designer, Mike Klymkowsky, who is closely involved in this project. While the scientific content of the
curriculum has been carefully thought through, the software artifact will cause many problems for students until it is subjected to thorough testing in naturalistic situations.

**Medical PBL**

Beginning students in the Medical School at Southern Illinois University have the option of taking a Problem-Based Learning (PBL) track for their first two years. Traditional educational assessments indicate that students opting for this are better prepared for the rest of their medical studies than those who attend lectures (Barrows, 1994). PBL students learn by working collaboratively in teams of five students and a tutor (facilitator) to investigate specific medical cases (problems). Koschmann has undertaken many studies of this approach to education, increasingly relying upon micro-ethnography and collaborating with trained micro-ethnographers. His experience has provided insight into the workings of collaborative learning in this particular successful setting, and has convinced us of the potential of micro-ethnography as a methodology for studying the role of artifacts in collaborative learning.

**Artifacts Seminar (http://www.cs.colorado.edu/~gerry/readings)**

The PI organized a seminar on artifact theory, primarily as a pilot project for this proposal. Core members of the project team met along with other faculty and graduate students from Communication, Education, Philosophy, and Computer Science. We reviewed theoretical texts on the nature of artifacts from cognitive science, CSCL, communication, cultural studies, psychology, philosophy, and social theory. We also held data sessions on episodes from the SIMROCKET tapes. Out-of-class discussions were held in WEBGUIDE and we conducted a SIMROCKET experiment mediated by WEBGUIDE.

**StateTheEssence (http://www.cs.colorado.edu/~gerry/projects/essence)**

STATETHEESSENCE is Web-based software developed by the PI to help middle school students develop their text summarization skills. It was used in an interdisciplinary four-year research project at a local Boulder public school. After undergoing considerable revision and refinement based on testing with students, the software was shown to improve text summarization, particularly in cases where the original text was somewhat difficult for the student to understand (Kintsch et al., 2000; Stahl & dePaula,
Evaluation of this software was conducted by means of controlled experiments and teacher ratings.

**JIME (http://www-jime.open.ac.uk)**

The web-based Journal of Interactive Media in Education, JIME, conducts group reviews of submitted articles online, and then includes an edited version of the review discourse with the published version. We are currently analyzing the online review discussions to draw conclusions about how the journal software and practices might be improved (Lenell & Stahl, 2001). This study provides us with some experience in analyzing online discourse, which will be important in the final stage of our proposed project.

### 8. Results from Prior Support

**Organizational Memory and Organizational Learning**

“Conceptual Frameworks and Computational Support for Organizational Memories and Organizational Learning (OMOL),” PIs: Gerhard Fischer, Gerry Stahl, Jonathan Ostwald, September 1997 – August 2000, $725,000, from NSF CSS Program #IRR-9711951.

This grant led to the current proposal’s focus on Web-based learning environments. The OMOL project started from a model of computer support for organizations as Domain-Oriented Design Environments (DODEs) in which both domain knowledge and local knowledge are stored in the form of artifact designs and associated design rationale. This CSCW model evolved into one of Collaborative Information Environments (CIEs), that emphasized the interactive, asynchronous, persistent discussion of concepts and issues within an organization. Gradually, interest in organizational learning aspects led to involvement in CSCL and the model of collaborative Knowledge-Building Environments (KBEs). A number of software prototypes were developed to explore the use of the Web as a communication and collaboration medium:

- **DYNACLASS**: A discussion forum for use in college courses. It features ties to DYNAGLOSS and SOURCES as well as email notification and specialized displays.

- **WEBGUIDE**: Differs from DYNACLASS in providing more control over rearrangement of notes; features computational perspectives.
• DYNAGLOSS: A system for defining technical terms and keywords and for debating the definitions and reviewing the history of debate; linked to DYNACLASS and SOURCES in that each term shows all the locations in these other systems where the term is explicitly referenced.

• SOURCES: A system for annotating bibliographical entries; uses terms from DYNAGLOSS as keywords.

• INFOMAP: An interface component for creating a graphical display of linked notes like a threaded discussion; providing convenient drag-and-drop functionality.

Work on this grant led to the focus on KBES as models of computer support for organizational memory and collaborative learning. In particular, it provided a number of different systems, each with useful functionality. As we tested and deployed these systems, we confronted serious issues of adoption and focused our concerns increasingly on socio-technical and social informatics issues: motivation, media competition, critical mass, social practices, seeding, management, re-seeding, convergence of ideas, peer-to-peer collaboration, deployment strategies. These issues led to a new research agenda (Stahl, 2000b) and this proposal.

Environmental Perspectives in a Middle School Classroom


This grant funded the initial implementation of WEBGUIDE as an integrated JAVA applet KBE supporting personal and group perspectives. It was a joint effort between the PI, a middle school teacher, and a research group at the National Oceanographic and Atmospheric Administration (NOAA) labs in Boulder. The teacher taught an environmental science class in which he wanted to spend the year having his students interview various adults and construct a set of contrasting perspectives (conservationist, regulatory, business, community) on a particular local environmental issue that the students had previously been involved in. WEBGUIDE was used by the students to collect notes on their interviews and to formulate personal and team perspectives on the issue. Results of this software trial were analyzed and presented at the AERA, CSCL, ICLS, CILT, WebNet, and Group conferences (Stahl, 1999c). These findings led to a number of revisions of WEBGUIDE, including the separation of the perspectives mechanism from the Web interface, and recognition of the need for software
architectures, standards, and components to support flexible rapid prototyping of KBEs.

**Interoperability among Knowledge-Building Environments**

“Interoperability Among Knowledge-Building Environments,” PI: Gerry Stahl, September 1999 – August 2000, $9,124, from NSF-funded Center for Innovative Learning Technology (CILT), Subcontract #17-000359 under NSF grant #EIA-9720384.

This was a seed grant whose purpose was to stimulate collaboration among KBE research groups. Part of the intention of the grant was to prepare a proposal for fuller funding, such as the present proposal. This grant resulted in a semester-long student project involving three graduate and three undergraduate students (one collaborating virtually from Germany using WEBGUIDE) creating an XML DTD that defines a data format for data imported from several different KBE prototypes and displayed in a Web browser using XSL. The grant supported a workshop organized by the PI, entitled “Collaborating on the Design and Assessment of KBEs in the 2000's” at CSCL '99 at Stanford. This workshop attracted over 60 participants and was preceded by an on-line discussion of 28 submitted position papers. This grant led to the emphasis on collaboration among KBE research groups and the need to develop and disseminate theory and methodology for developing educational artifacts, as proposed here.

**Incorporating Automated Text Evaluation into Collaborative Software Environments**

“Allowing Learners to be Articulate: Incorporating Automated Text Evaluation into Collaborative Software Environments,” PIs: Walter Kintsch, Gerhard Fischer, Thomas Landauer (Stahl served as co-PI and primary software developer), calendar 1997-2000, $1,400,000 for four-years, from the McDonnell Foundation’s Cognitive Science in Education (CSEP) program.

This grant supported the design, implementation, and testing of STATETHEESSENCE, Web-based software to teach middle school students summarization skills (Kintsch et al., 2000; Stahl & dePaula, 2001; Steinhart, 2000).
New Media to Support Collaborative Knowledge Building


The grant is to test WEBGUIDE in the Artifacts Seminar and to make a number of technical improvements to WEBGUIDE’s functionality.

9. Contributions to ROLE Goals and Potential Impact

Recent research on learning and on technology in education – as surveyed in the Report to the President (Panel on Educational Technology, 1997) and in How People Learn (Bransford et al., 1999) – stresses the potential of innovative constructivist educational approaches to foster deep understanding. The latter document, for instance, concludes that computer technology “has great potential to enhance student achievement and teacher learning, but only if it is used appropriately” (Ch. 9).

Pursuing the Potential of Computer-Supported Collaborative Learning (CSCL)

Our project proposes to investigate at a detailed level the key learning processes in a computer-supported collaborative knowledge-building environment. We believe that collaborative learning has a great potential to foster deep knowledge building when it brings together the perspectives of multiple students in a productive way. However, this requires a more detailed understanding of how collaborative knowledge-building processes work. We further anticipate that computer support has the potential to facilitate collaboration by removing communication limitations and by helping to manage the complexity of ideas and interactions. However, this requires carefully designed knowledge-management software applications tuned to the needs of collaborative learning. This project will investigate the potential of CSCL through focused inquiry using micro-ethnographic methods to observe how learning takes place in collaborative settings under conditions of computer support.
A Methodology for Assessing Computer Support for Collaborative Learning

We will develop a methodology for assessing the role of computational cognitive artifacts in supporting collaborative knowledge building. We will start with micro-ethnography as a methodology for analyzing the construction of social organization in small group communicative interactions. We will adapt this methodology to address the concerns of human-computer interaction. This will provide future researchers and designers with an approach for studying CSCL at a level of detail that can usefully drive iterative software design. Our approach will make visible the learning that is taking place within small groups of students engaged in computer-supported tasks, thereby indicating what is needed to help students learn how to take advantage of the computational artifacts and where the design of the artifacts is effective or problematic.

Increased Understanding of the Role of Computational Cognitive Artifacts in Learning and Education

Our project will increase our understanding of the social origins and maintenance of knowledge, how knowledge is embodied in artifacts, how the artifacts transmit this knowledge, and how people learn to use the artifacts to make cognitive use of the embedded knowledge. In complex learning systems like VIRTUALBIOLOGYLAB, such knowledge is presented in many forms: explicit textual knowledge like background information and procedural instructions, animations and other representations of physical processes, simulations that generate data for analysis, interactive tasks, mini-tests of student understanding, interactive tools for manipulating data, and lab results that require guided interpretation. We will study how students in small groups successfully and problematically collaboratively construct these various forms of knowledge through interaction with the software and with each other. This will increase our theoretical understanding as CSCL researchers of the role of computational cognitive artifacts in learning and education.

Impact on the CSCL Research Community

There is a vigorous and growing CSCL research community in the United States and globally, despite rather limited funding opportunities here beyond ROLE. Last year over 700 people attended CSCL ’99 and this year the first European CSCL conference will take place. In January 2002, CSCL will be hosted at the University of Colorado, with all of this project’s staff actively involved. The focus of the conference will be on new methodologies and
deepened theoretical frameworks. We anticipate that our project will provide an organizational basis for establishing at CSCL 2002 on-going global research collaborations. Our project will provide an example of a systematic attempt to apply a new methodology within the field and to elaborate a coherent theoretical framework that addresses core issues of CSCL.

**ROLE Quadrant 2 and 3**

While ROLE projects related to neuroscience may focus quite literally inside the head of an individual learner, this project will look outside at the social interactions through which knowledge is constructed and shared – and at the same time evidenced – in collaborative educational settings. The project is situated in ROLE’s quadrant 2 because it builds bridges from the cognitive sciences to research on learning. Through micro-ethnographic studies of educational environments, it undertakes fundamental research on behavioral, cognitive, affective, and social aspects of human learning as mediated by artifacts. It is also situated in quadrant 3 because it is building a stronger research base to support educational approaches (e.g., scaffolded collaborative small groups), curriculum materials (VIRTUALBIOLOGYLAB), and technological tools (WEBGUIDE) to facilitate the learning process. In particular, the project builds on diverse cognitive theories of the role of artifacts and on methods of micro-ethnography in order to develop and refine new education research and evaluation methods for analyzing the role of computational cognitive artifacts in collaborative learning and education – and investigates these theories and methods in formal science learning settings in middle school, high school and college.

**References**


Barrows, H. (1994) Practice-based Learning: Problem-Based Learning Applied to Medical Education, SIU School of Medicine, Springfield, IL.


Technologically Supported Environments, Springer Verlag, Heidelberg, Germany.


Lenell, E. & Stahl, G. (2001) Evaluating affordance short-circuits by reviewers and authors participating in on-line journal reviews (submitted), In: Proceedings of European Computer Supported Collaborative Learning (ECSCL '01), Maastricht, NL.


Stahl, G. (1999a) POW! Perspectives on the Web, In: Proceedings of WebNet World Conference on the WWW and Internet (WebNet '99), Honolulu, HA. Available at:


Streeck, J. (1983) Social Order in Child Communication: A Study in Microethnography, Benjamins, Amsterdam, NL.
Perspectives on Collaborative Knowledge Building

**Knowledge Building**: This project proposes to develop an approach to information management grounded in an innovative theory of learning and collaboration as knowledge building. The theory is oriented toward guiding development of technology that can better fulfill growing societal requirements for Web-based support of groups. It offers an analysis of collaborative knowledge-building activities and of Knowledge-Building Environments (KBEs) to support these activities. To do this, it brings together and synthesizes approaches and concepts from situated learning, activity theory, hermeneutics, distributed cognition, and related theories from philosophy, social science, education, and computer science.

**Knowledge-Building Environments**: The proposed work centers on a new way of using computers to support collaboration by integrating information management support for various activities of knowledge building. Such KBE software prototypes go beyond superficial discussion or chat and beyond choices and decisions among fixed options to support the co-construction of deep knowledge, innovation, and shared understanding. They support learning, working, and innovation over time within groups that may be physically distributed.

**Perspectives**: A key innovation for KBEs in this project is the modeling of the interplay between individuals and the group in collaborative settings. This is done by providing personal and group computational perspectives: virtual workspaces whose contents are interrelated by automated inheritance mechanisms. Computational perspectives provide a new, dynamic, personalized form of on-line information management that supports the fundamental structure of collaboration. They help to manage a shared information space so that participants view information relevant to themselves and can process (edit, rearrange, reconceptualize) that information without affecting anyone else’s personal perspective. Then, through supported negotiation activities, information is migrated to subgroup and group perspectives, where it represents shared knowledge.
Research and Training Focus: The project will directly employ four students and will involve many more in seminars and class projects. Because KBE software is essentially a new form of learning technology, education students as well as computer science and other students will be involved in designing, developing, deploying, and assessing the software. This will help to create an interdisciplinary educational technology research focus that the PI and his colleagues have already begun to foster.

Software Development: The proposed work will build an infrastructure for local, national, and international collaboration on KBE software development, providing a more reliable basis for assembling KBE prototypes customized to particular deployment sites. An open source perspectives server will be released, allowing researchers to develop KBE interface components that simply call this server for database access and dynamic computation of perspective contents. A standard for data interchange with the server and for interoperability among KBE systems will facilitate a component architecture and the use of shared tools to assess KBE usage.

Study Sites: The project will assess KBEs with perspectives in realistic study sites: an academic research group, a collaborative learning seminar, a corporate training setting, and an industrial design group. Quantitative analysis of captured textual contents will be compared with results from non-integrated threaded discussion systems and other groupware. Qualitative analysis of surveys and field notes will investigate issues of deployment, adoption, social practice, utility, and effectiveness.

Project Impacts: The proposed project should result in progress in the development and assessment of KBEs, an emerging form of software with a potential to significantly extend human cognition by supporting collaborative knowledge-building activities and by providing persistent external memory of what took place during the collaboration. The release of a perspectives server with its associated standards will provide a concrete basis for catalyzing local, national, and international collaboration among KBE researchers. This will promote KBE research as an important new research focus.

Project Description

This project is guided by a theory of collaborative Knowledge-Building Environments (KBEs) that we are developing. This theory proposes the following principles:
• Collaborative knowledge building is a particular view of group learning that focuses on a range of activities that take place within communities, as opposed to focusing on learning as the transmission of bits of information to individual learners.

• Collaborative knowledge building takes place largely through the interaction among people with different understandings from multiple personal and group perspectives.

• Such knowledge building within groups can be helped by appropriately designed information technology (IT) that supports various knowledge-building activities and supports interaction among alternative perspectives.

The form of IT that we are interested in – collaborative Knowledge-Building Environments – represents a distinctive approach that overlaps related work in Computer-Supported Collaborative Learning (CSCL) and Computer-Supported Cooperative Work (CSCW). IT support for learning is traditionally oriented toward the transmission of information to individual students. Even where it is based on a view of student construction of knowledge, as with Intelligent Tutoring Systems (ITS) for algebra or physics, the goal is measured by testing the incorporation of pre-defined content or methods into the individual’s understanding (Wenger, 1987). A more student-centered, constructivist approach is taken by Interactive Learning Environments (ILE), which might, for instance, allow students to create ecologies in SimLife to learn biology, or programs in Turtle Logo to explore math concepts (Papert, 1980). In contrast, a KBE primarily supports the group process and leaves matters of content up to the participants (which may include a teacher who raises particular content issues and helps maintain focus). In this way, it applies CSCW approaches to CSCL. A review of CSCW technology for groups (Kraemer & Pinsonneault, 1990) distinguishes group communication support systems (GCSSs) from decision support (GDSSs). GCSSs are specific communication media like email and video-conferencing. In providing computational tools for group decision making, GDSSs tend to support isolated, focused activities that integrate products of individual work. In contrast, a KBE aims to support a broad spectrum of knowledge-building activities – both individual and group – in a more seamless fashion. It supports the construction of areas of knowledge through group inquiry over extended periods of time. It also supports the interplay of individual and group more comprehensively, through integrated mechanisms of "computational perspectives" and negotiation that treat the group as more than just the sum of the individuals.
Assessments of CSCL and CSCW systems have defined a number of key issues for evaluating the problems and successes of such systems. For instance, in simple threaded discussion forums common problems include: short threads (a tendency for discussions to die quickly), low participation (lack of motivation to participate), few cross-references (little convergence of ideas), and superficial content (minimal depth of investigation) (dePaula, 1998; Guzdial & Turns, 2000; Hewitt & Teplovs, 1999). On the other hand, GDSSs and GCSSs attempt to decrease communication barriers within the group, while increasing task-oriented focus, depth of analysis, and decision quality (Connolly, 1997; Kraemer & Pinsonneault, 1990). Social informatics studies have raised additional issues of software deployment and adoption in addition to questions of usability and utility (Kling, 1999). These are some of the dimensions along which KBEs must be assessed within realistic learning and working social contexts.

To date, the PI and his colleagues have begun to develop KBE theory in conjunction with Web-based KBE prototypes that support many of the activities described in the theory and that have been tested informally in collaborative learning classrooms. In particular, computational support for personal and group perspectives has been developed and tried out. Support for computational perspectives was explored in the PI’s dissertation (Stahl, 1993a) and has since been refined and adapted to the Web (Stahl, 1999a). This work has been described in relevant CSCL and CSCW conferences (see references by Stahl). The proposed project will build on existing concepts and prototypes, extending them substantially by: implementing a technical infrastructure to support data interoperability, integration of functionalities, and rapid prototyping; deploying customized KBEs in specific study sites; observing the social impacts of these IT systems in the work settings; revising the theory based on empirical findings; and fostering a community of researchers working on IT support for knowledge building in workgroups.

1. PREPARATION FOR PROPOSED WORK UNDER PRIOR NSF SUPPORT

1.1. Organizational Memory and Organizational Learning (CSS)

“Conceptual Frameworks and Computational Support for Organizational Memories and Organizational Learning (OMOL),” PIs: Gerhard Fischer, Gerry Stahl, Jonathan Ostwald, September 1997 – August 2000, $725,000, from NSF CSS Program #IRR-9711951.
This grant prepared much of the background for the proposed work. The OMOL project started from a model of computer support for organizations as Domain-Oriented Design Environments (DODEs) in which both domain knowledge and local knowledge are stored in the form of artifact designs and associated design rationale (Fischer, 1994). This CSCW model evolved into one of Collaborative Information Environments (CIEs), that emphasized the interactive, asynchronous, persistent discussion of concepts and issues within an organization (Stahl, 1998; Stahl, 2000a). Gradually, interest in organizational learning aspects led to involvement in CSCL and the model of collaborative Knowledge-Building Environments (KBEs) (Fischer et al., 1999). A number of software prototypes were developed to explore the use of the Web as a communication and collaboration medium. Of these, the most important for the proposed work are the following:

- **DynaClass**: A discussion forum for use in college courses. It features ties to DynaGloss and Sources as well as email notification and specialized displays (Ostwald, 1999).

- **WebGuide**: Differs from DynaClass in providing more control over rearrangement of notes; features computational Perspectives (Stahl & dePaula, 1998; Stahl et al., 1998).

- **DynaGloss**: A system for defining technical terms and keywords and for debating the definitions and reviewing the history of debate; linked to DynaClass and Sources in that each term shows all the locations in these other systems where the term is explicitly referenced.

- **Sources**: A system for annotating bibliographical entries; uses terms from DynaGloss as keywords.

- **InfoMap**: An interface component for creating a graphical display of linked notes like a threaded discussion; providing convenient drag-and-drop functionality.

Work on this grant led to the focus on KBEs as models of computer support for organizational memory and organizational learning. In particular, it provided a number of different systems, each with useful functionality, and brought home the need to define component standards so the functionalities can be combined more flexibly. As we tested and deployed these systems, we confronted serious issues of adoption and focused our concerns increasingly on socio-technical and social informatics (Kling, 1999) issues: motivation, media competition, critical mass, social practices, seeding, management, re-seeding, convergence of ideas, peer-to-peer collaboration, deployment strategies. These issues led to a new research agenda (Stahl, 1999b) and this proposal.
1.2. WebGuide and Environmental Perspectives (NOAA)


This grant funded the initial implementation of WebGuide as an integrated Java applet KBE supporting personal and group Perspectives. It was a joint effort between the PI, a middle school teacher, and a research group at the National Oceanographic and Atmospheric Administration (NOAA) labs in Boulder. The teacher taught an environmental science class in which he wanted to spend the year having his students interview various adults and construct a set of contrasting perspectives (conservationist, regulatory, business, community) on a particular local environmental issue that the students had previously been involved in. WebGuide was used by the students to collect notes on their interviews and to formulate personal and team perspectives on the issue. Results of this software trial were analyzed and presented at conferences (Stahl, 1999a; 1999b; 1999c; Stahl & Herrmann, 1999). These findings led to a number of revisions of WebGuide, including the separation of the Perspectives mechanism from the Web interface, and recognition of the need for software architectures, standards, and components to support flexible rapid prototyping of KBEs.

1.3. Collaboration in KBEs (CILT)

“Interoperability Among Knowledge-Building Environments,” PI: Gerry Stahl, September 1999 – August 2000, $9,124.21, from NSF-funded Center for Innovative Learning Technology (CILT), Subcontract #17-000359 under NSF grant #EIA-9720384.

This is a current seed grant whose purpose is to stimulate collaboration among KBE research groups. Part of the intention of the grant was to prepare a proposal for fuller funding, such as the present proposal and its currently pending complementary NSF proposals for “IT Support for Knowledge Building in Workgroups” and “ITR/EWF: Collaborative Research on Knowledge-Building Environments: Growing a National and International Research Community for Distance Learning Information Technology.” This grant has already resulted in a semester-long student project involving three graduate and three undergraduate students (one collaborating virtually from Germany) creating an XML DTD that defines a data format for data imported from several different KBE prototypes and displayed in a Web browser using XSL. The grant supported a workshop entitled “Collaborating on the Design and Assessment of KBEs in the 2000's” at CSCL ’99 at Stanford. This workshop attracted over 60
participants and was preceded by an on-line discussion of 28 submitted position papers. This grant has led to the emphasis on collaboration among KBE research groups and the need to put into place some of the technical and social conditions for such collaboration (Stahl, 1999a), as proposed here.

2. OVERVIEW OF PROPOSED WORK

2.1. Theory of Collaborative Knowledge-Building Environments

Collaborative Knowledge Building

Information Technology (IT) is a broad field that can be conceptualized in various ways. Traditionally, the computer was thought of as a medium for storing and delivering data, that can then be used by people in their work. More recently, the computer (especially with the Web) has become a medium of communication, through which people share information and knowledge. This communication can take a variety of forms. In simple forms of e-commerce or on-line voting, people submit their decisions about a fixed list of choices. In chat and most email, people exchange greetings and opinions, generally without changing those opinions. Many systems in recent years have tried to support a particular form of communication or social interaction like brainstorming or decision-making – often with very positive results (Connolly, 1997; Vogel et al., 1987). We are interested in a distinct but broader process of communication which we term collaborative knowledge building. Here, groups of people construct new knowledge through interaction of their ideas and perspectives, usually eventually preserved in documents or other artifacts.

Our theory of collaborative knowledge building (Fischer et al., 1993/1998; Stahl, 1975; 1993a; 1993b; 1999c; 2000a; 2000b; Stahl & Herrmann, 1999) proposes a concept we call the synergistic moment; we intend to investigate the validity of this concept in the proposed project. The synergistic moment is the critical point during collaboration in which a group constructs meaning that transcends what any participant may have “in mind.” The shared understanding that is generated in this process is a subtle phenomenon: It does not mean that everyone is in complete agreement or even that each individual has the same internal cognitive representations of
what is discussed. Rather, it means that a certain group view has been expressed. The unit of analysis for describing this is the group, and is manifested in the group's discourse. Individuals may agree to disagree with the group understanding, and careful investigation may reveal that individual understandings differ from the group's view (Hatano & Inagaki, 1991). The intersubjective "sharing" is not a correspondence or overlapping of individuals' mental content, but a coordination or interaction of their participation in joint socio-cultural activity (Matusov, 1996). The synergistic moment is an emergent property of the group dialog as a cacophony of voices (Bakhtin, 1986). It could easily pass unnoticed as a magical fount of creativity; to more deeply understand it likely requires "thick description" (Geertz, 1973) and detailed interaction/discourse analysis (Jordan & Henderson, 1995), and therefore presupposes that the interaction was captured in some medium. Fortunately, the literature on CSCL contains a number of incisive analyses (Roschelle, 1998) of the synergistic moment, although they do not highlight it as such.

The synergistic moment is a result of perspective-sharing (Boland & Tenkasi, 1995), but at the group rather than the individual level. It overcomes the problem pointed out by Feltovich et al. (Feltovich et al., 1996), that any one perspective may limit the ability to comprehend creatively the complexity of a topic under discussion. What typically happens is that one person makes a statement from her personal perspective; someone else interprets that statement from his own perspective and responds accordingly; others continue this process so that the discourse consists implicitly of reinterpretations from various perspectives. The drive to establish intersubjectivity and shared knowledge is powered by socio-cognitive conflict and contention among perspectives according to studies by Piaget and his followers (Perret-Clermont & Schubauer-Leoni, 1981). The dialog proceeds through sequential turn-taking and attempts to repair "misunderstandings" as understood from particular perspectives and reinterpreted from others. Thanks to the human drive to impose coherent social meaning structures (Geertz, 1973), a synergistic group understanding emerges. This shared understanding can play a central role in the further activity of the group and can be more or less adopted by individuals into their personal perspectives. Although the synergistic moment seems to the participants to emerge spontaneously, it can be understood as the result of many identifiable knowledge-building activities, as represented in our model (below).
Perspectives in Knowledge Building

According to hermeneutics – the philosophy of interpretation – human understanding is fundamentally perspectival. We construct knowledge from our situated perspective in the world: our historical position, cultural tools, and practical interests (Gadamer, 1960/1988; Heidegger, 1927/1996; Stahl, 1975). Computational support for knowledge building can represent our interpretive perspectives with computational Perspectives (Boland & Tenkasi, 1995; Nygaard & Sørgaard, 1987; Winograd & Flores, 1986). (In this proposal, Perspective–with-a-capital-P will refer to the proposed computational mechanism that mirrors human interpretive perspectives–with-a-lower-case-p.) In this sense, Knowledge-Building Environments (KBEs) with computational Perspectives are designed to support the essential structure of collaboration. A key hypothesis of the proposed work is that KBEs benefit from an approach that represents the perspectival nature of collaboration. A goal of the project is to facilitate the incorporation of a computational Perspectives mechanism in KBEs – both in our own prototypes and in the work of other KBE research groups around the world.

Computational Perspectives have been explored by the PI in a number of software prototypes, in his dissertation system, and in his theoretical publications (Stahl, 1993a; 1993b; 1995; 1998; Stahl & Herrmann, 1999; Stahl et al., 1995). In a single-user system, computational Perspectives may correspond to different domains or professional viewpoints on a design problem, such as electrical, plumbing, structural, and heating concerns in architecture (Fischer et al., 1993; 1993/1998). In a KBE to support collaboration, computational Perspectives typically provide personal or group workspaces for the development of different sets of ideas. In this way, they can model the relationships among the various personal and group interpretive perspectives at work in the construction of collaborative knowledge.

We hypothesize that computational Perspectives can support the synergistic moment in collaborative knowledge building by providing the necessary contact among different personal Perspectives, allowing them to interact, and then locating the results in a group Perspective. By situating the traditionally ephemeral synergistic moment within an explicit structure of computational Perspectives and by doing so in a persistent way, a KBE provides new opportunities for group self-reflection.

An important complement to Perspectives is negotiation. Negotiation is a process through which divergent personal perspectives converge on a collaborative shared understanding. When Perspectives and negotiation are effectively “intertwined” in a KBE, they compensate for each other’s
potential problems: Negotiation converges ideas so that everyone can benefit from the ideas of other perspectives, while personal Perspectives allow people to work on their own views while potentially time-consuming negotiations are underway (Stahl & Herrmann, 1999).

For instance, when WebGuide – a KBE with computational Perspectives implemented by the PI – was used in a middle school environmental science classroom, students each had their own personal Perspective in which to develop their own responses to questions posed by the teacher. The teacher’s questions to the whole class were posed in the class’ group Perspective. From there they were automatically inherited into the team Perspectives. The content of the team Perspectives was, in turn, inherited into the personal Perspectives of team members. Gradually, students migrated their ideas to team Perspectives that represented either conservationist, governmental, corporate, or citizen perspectives on the ecological controversy – depending on which perspective team the student was part of. Then they could work with the ideas of their team-mates and negotiate their team position. In the end, the different teams negotiated to spell out agreements and disagreements (Stahl, 1999c).

The analysis of the synergistic moment suggests that negotiation need not take the explicit, rationalist forms typical of GDSSs, such as voting. Group results may emerge naturally out of the intertwining of Perspectives in group discussion. A challenge of the proposed work will be to develop software support for capturing such results and migrating them un-intrusively to group Perspectives.

The Potential of IT Support for Knowledge Building

IT support has the potential of transforming the activities underlying the synergistic moment. For one thing, it would make those activities publicly accessible. The group could then reflect upon the emergence of its shared understanding by looking over the persistent record of its dialog. Such reflection might prove especially useful in contentious situations or for newcomers who were not part of the original dialog and are motivated to re-open the issue – as illustrated by Matusov (1996). Furthermore, computer support of perspectives could make explicit the interplay of different personal Perspectives and the migration of ideas and their interpretations between personal and group Perspectives. Ironically, perhaps, the “asynchronous” medium of the Web would allow group members to interact simultaneously – without waiting for sequential turns – thereby overcoming what Peters (1998) characterizes as “the hardest argument against democracy: the ability of only one person to speak and be heard at a time”
Proposals for Research

(p. 261). Of course, as we have already discovered with the Web in general, the increased flood of ideas raises complex information management issues. We do not yet understand the full social impact of the envisioned KBEs – we will only know how they are used once they have been implemented, deployed in naturalistic settings, and observed.

Theories of human cognitive development emphasize the important role of external memories to extend short-term and long-term human memory (Donald, 1991; Norman, 1993). They also stress that individual cognition is a social product, highly mediated by social symbol systems, cultural artifacts, processes of structuration, and group collaboration (Bourdieu, 1972/1995; Geertz, 1973; Giddens, 1984; Hutchins, 1996; Vygotsky, 1930/1978). This suggests that computer support for collaboration has the potential to significantly advance the power of human cognition. In addition to maintaining a persistent external memory, IT can help people to be more reflective and creative – as has been demonstrated in computer support for brainstorming and decision-making (Connolly, 1997; Vogel et al., 1987). However, as our research to date indicates, despite the fact that the Web seems to offer a promising technological base for such a development, computer-supported collaboration is a complex process that requires a sophisticated body of knowledge that we are just beginning to assemble. Moreover, the potential is beyond the reach of any single research group.

We believe that IT support for collaborative knowledge building has not yet been developed to near its potential. KBE research has been carried on for over a decade now, starting with the CSILE system and continuing with KIE, CoVis, etc. (Cuthbert, 1999; Pea, 1993; Scardamalia & Bereiter, 1991; 1996). Recently, commercial systems like KnowledgeForum, WebCT, and LearningSpace are catching on. However, as yet there has been no systematic attempt to support the variety of activities that are involved in knowledge building. There is no general theory of collaborative knowledge building as a social process. Existing research tends to target specific contexts like middle school science with specialized closed systems, rather than developing interchangeable, open source components that can be applied in a full range of contexts. While networks of KBE researchers are coming together in other countries, there is little organized effort to collaborate in the US. The proposed project aims to change this situation. However, collaboration across institutions cannot be started by just wishing for it. This project tries to put some of the necessary conditions in place by developing technical infrastructure (standards, a Perspectives Server, interface components) and initial results that can be used to stimulate discussion and collaboration among KBE researchers locally, nationally, and internationally. Interoperability and collaboration will allow isolated
advances to be exchanged, new functionalities to be shared, and test data to be compared.

**A Model of Collaborative Knowledge Building**

One approach to better understanding how to design computer support for collaborative knowledge building in social settings is to conceptualize the various constituent activities involved in individual and social knowledge building. The diagram below from (Stahl, 2000b) provides a starting point for this, combining aspects of activity theory, situated learning, hermeneutic philosophy, and distributed cognition theory (Chaiklin & Lave, 1993; Cole, 1996; Engeström et al., 1999; Gadamer, 1960/1988; Hutchins, 1996; Lave & Wenger, 1991; Nardi, 1996).

The idea of this diagram is that knowledge building can proceed through many different activities. The sequential structure of the model is only illustrative of an ideal conceptualization. We understand that these activities complexly overlap in practice. The possible relationships among the individual activities – and particularly the interactions between the personal and social – can be complex and varied. The purpose of the diagram is to suggest a number of distinct activities that could be supported by a KBE with multiple functionality. The sequential labeling of these activities corresponds to proposed KBE components listed in Table 1 below, and it is not intended to imply a necessary order to the activities.

Figure 1. A model of personal understanding and social knowledge building.

A set of seminal books and articles in Computer-Supported Collaborative Learning (CSCL) has formulated a view of learning as a social process of collaborative knowledge building within communities of practice (Brown &
Campione, 1994; Brown & Duguid, 1991; Lave, 1991; Lave & Wenger, 1991; Pea, 1993; Scardamalia & Bereiter, 1996; Wenger, 1998). However, these texts do not make the set of cognitive and social activities that underlie such a view explicit in the manner attempted in our KBE theory.

Starting in the lower left corner, Figure 1 shows a cycle of personal understanding. The rest of the diagram depicts how personal beliefs can be articulated in language and become part of social interaction. Note that the results of social knowledge building eventually feed into personal understanding, providing the evolving toolkit of culturally-based individual cognitive capabilities. The depicted knowledge-building activities are discussed briefly below in the context of proposed computer support.

**IT Support for Knowledge-Building Activities**

Each of the activities of social knowledge building pictured in Figure 1 can be supported computationally. Table 1 lists an illustrative form of support for each. It also lists corresponding prototypes that we have developed. Support for each activity is briefly discussed following the table.

Table 1. Forms of computer support for knowledge building activities.

<table>
<thead>
<tr>
<th>Knowledge-building activities</th>
<th>Forms of computer support</th>
<th>Prototype systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>a articulate in words</td>
<td>articulation editor</td>
<td>DynaClass</td>
</tr>
<tr>
<td>b public statements</td>
<td>personal Perspective</td>
<td>WebGuide</td>
</tr>
<tr>
<td>c other people's public statements</td>
<td>comparison Perspective</td>
<td>WebGuide</td>
</tr>
<tr>
<td>d discuss alternatives</td>
<td>discussion forum</td>
<td>DynaClass</td>
</tr>
<tr>
<td>e argumentation &amp; rationale</td>
<td>argumentation graph</td>
<td>InfoMap</td>
</tr>
<tr>
<td>f clarify meanings</td>
<td>glossary discussion</td>
<td>DynaGloss</td>
</tr>
<tr>
<td>g shared understanding</td>
<td>glossary</td>
<td>DynaGloss</td>
</tr>
<tr>
<td>h negotiate perspectives</td>
<td>negotiation support</td>
<td>WebGuide</td>
</tr>
<tr>
<td>i collaborative knowledge</td>
<td>group Perspective</td>
<td>WebGuide</td>
</tr>
<tr>
<td>j formalize and objectify</td>
<td>bibliography discussion</td>
<td>Sources</td>
</tr>
<tr>
<td>k cultural artifacts and representations</td>
<td>bibliography or other community repository</td>
<td>Sources</td>
</tr>
</tbody>
</table>

a. Computer support should facilitate the process of articulating ideas and preserving them in convenient forms. Most KBEs, including discussion forums like DynaClass, provide an editor for articulating ideas. Some KBEs have tried to introduce procedural facilitation, scaffolding, or prompting to encourage someone to articulate an appropriate expression (Slotta & Linn, 2000). Other approaches would be to provide an outline editor or a brainstorming area.
b. Public statements by one person confront those of other people. Computer support can represent the different perspectives from which these statements emerge. Perspectives are more general than representations of individuals themselves, because one person can offer statements from multiple perspectives and several people can agree on a common perspective. Perspectives can be related to one another, for instance deriving from a common perspective that they share. Computational representations of perspectives in a KBE like WebGuide (Stahl, 2000a) make explicit the important relationships among personal and group perspectives, as well as providing means for individuals and collaborative teams to articulate their own perspectives.

c. A KBE with support for Perspectives should provide comparison Perspectives, in which one can view and contrast alternative Perspectives and adopt or adapt ideas from other people's Perspectives. Comparison Perspectives in WebGuide aggregate ideas from various individual and/or group Perspectives and allow for comparison of them (Boland & Tenkasi, 1995; Stahl, 1999c). Other systems like D3E (Sumner & Buckingham Shum, 1998b) facilitate commentary on documents by other people, such as reviews of journal articles.

d. The most common element in current KBEs is the discussion forum. This is an asynchronous, interactive communication system like DynaClass that allows people to respond to notes posted by one another. Typically, there is a thread of responses to entered notes, with a tree of divergent opinions. A KBE should go beyond superficial undirected discussion to converge on shared understandings (dePaula, 1998; Guzdial & Turns, 2000; Hewitt & Teplovs, 1999).

e. Although every note in a discussion forum is a response to another note, the discussion may have a more complex implicit structure. One note might argue for or against another or provide evidence to back up the claim of another note, for instance. Such an argumentation structure can be made explicit and formalized in a representation of the argumentation graph. A component like InfoMap that displays the structure of notes graphically can contribute to participants' meta-level comprehension of their knowledge-building activity, pointing out where additional evidence is needed or where alternatives have not been explored (Buckingham Shum & Hammond, 1994; Donath et al., 1999; Suthers, 1999).

f. An important requirement for constructing group knowledge is the establishment of shared understanding. This can be fostered by clarifying the meaning of important terms used in various competing
Proposals for Research

claims. A glossary discussion can make explicit how different participants understand the terms they use, as in DynaGloss or DocReview (Hendricksen, 1999).

g. The glossary discussion should result in a group glossary of the agreed upon definitions of important terms. Such a glossary already represents a form of group knowledge. The glossary is, of course, subject to future debate and emendation; it may make sense to define the glossary as a particular display of information from the glossary discussion (Stahl & Herrmann, 1999).

h. Perhaps the most delicate phase of knowledge building is negotiation. Computer support of negotiation tends by nature to make explicit the factors entering into the negotiation process. This can be extremely harmful to the subtle processes of persuasion if not done sensitively. On the other hand, negotiation is critical to helping multiple perspectives to converge on shared knowledge. Computer support can provide a useful tool—as long as it is carefully integrated with other social activities that allow for implicit, culturally established interpersonal interactions (Stahl, 1999b). Group Decision Support Systems (GDSSs) have traditionally been independent systems, not integrated with the broader context of knowledge building (Kraemer & Pinsonneault, 1990; Vogel et al., 1987).

i. The accumulation of negotiated shared knowledge results in the establishment of a group perspective. Like the alternative individual and team (or subgroup) perspectives, the group perspective may be represented in a KBE. In WebGuide, the content of the group Perspective is inherited into the individual and team Perspectives, because it has been accepted by the group. Individuals can then build on this shared knowledge within their own Perspective and even begin to critique it and start the whole cycle over (Stahl, 1999a).

j. Shared knowledge can be further formalized. It can be represented in another symbolic system or combined into a more comprehensive system of knowledge (Stahl, 1999c). For instance, in academic research knowledge is incorporated in new classroom lectures, conference presentations, journal articles, and books. The discussion of knowledge that has been compiled into publications can be carried out in a bibliography discussion component of a KBE.

k. Finally, representations of the new shared knowledge in publications and other cultural artifacts are themselves accepted as part of the established paradigm. Although still subject to occasional criticism,
ideas in this form more generally provide part of the accepted base for building future knowledge. In academic circles, an annotated bibliography like Sources might provide a useful KBE component to support this knowledge building activity (Sumner & Buckingham Shum, 1998a).

A KBE goes beyond a single-purpose system – like a simple discussion forum – and supports more than one collaborative knowledge-building activity (Muukkonen et al., 1999). It retains a record of the knowledge that was incrementally collected – unlike common chat, newsgroup, and listserv systems that erase contributions after a short period of time. We hypothesize that it should help people to express their beliefs, to discuss them with others, to differentiate their own perspectives and adopt those of other people, clarify disagreements or misunderstandings, critique and explicate claims, negotiate shared understandings or agreements, and formulate knowledge in a lasting representation. Because KBEs are computational, they can provide facilities like searching, browsing, filtering, tailoring, and linking in order to group related ideas together automatically. KBEs can interface with other agents and software utilities – for instance sending emails to notify collaborators when important knowledge-building events have taken place (McLean, 1999). They can also dynamically format sets of notes in convenient displays for different purposes.

2.2. An Infrastructure for KBEs

Computational Perspectives

Computational Perspectives provide a new, dynamic, personalized form of on-line information management (Stahl, 1995). A Perspective defines an electronic workspace in which a person or group can develop ideas and manage information that belongs together – for instance because it represents the beliefs and viewpoint of a particular person, group, domain, or intellectual position. Perspectives structure a shared information space so that special coherent views can be built up and displayed. Although the mechanism of computational Perspectives is very general and flexible, the simplest way to use it in a small group is to define a personal Perspective for each member, one team Perspective for agreed upon ideas, and a comparison Perspective that collects the ideas from all the personal Perspectives.

The design philosophy behind computational Perspectives as implemented by the PI in WebGuide is that users have complete control over the content
in their personal Perspectives. Thus, if my personal Perspective inherits conflicting ideas from different team Perspectives that I belong to, I can delete, edit, and rearrange those ideas at will. Other users can view the contents of my personal Perspective (except for content that I have designated as private) and they can copy items, link to them, initiate public discussions of them, and propose them for incorporation in team Perspectives – but none of this affects how the content of my Perspective is displayed to me. This allows me to build my own Perspective on the topics that are under consideration by the group. I can see what knowledge others are building, incorporate that knowledge into my Perspective, or join in with others to share, discuss, and negotiate. The same design philosophy applies of course to team Perspectives: team members jointly (through negotiation processes) have complete control over the content of their team Perspective.

Inheritance is a central defining mechanism of computational Perspectives as used in this proposal. The ability to define arbitrarily complex networks of Perspectives with multiple layers of sub-groups between the group Perspective and the individual personal Perspectives, and to have the automatic inheritance of content through the network distinguishes this approach from all other systems of “views” and “perspectives.” Inheritance in this sense is not class inheritance, but “content inheritance.” A given Perspective can inherit content from multiple other Perspectives. This content is aggregated (logical union) in the given Perspective, where it can be over-ridden with edits, deletions, rearrangements, and additions. The inheritance mechanism is derived from efficient approaches explored in hypermedia, including “delta memory” and “transclusion” (Boborow & Goldstein, 1980; McCall et al., 1990; Mittal et al., 1986; Nelson, 1981; Nelson, 1995). For a discussion of related work, see (Stahl & Herrmann, 1999).

Because new Perspectives can be defined (either in advance or during system use) to inherit from any (non-cyclical) other Perspectives, it is generally useful to define “comparison Perspectives” that aggregate the ideas from team members, including those ideas that have not been agreed upon and migrated to the team Perspective. This is handy for keeping an eye on what one’s fellow team members are thinking. Typically, we have set up the inheritance network of Perspectives to have a diamond-shaped profile, diverging out from the total group Perspective via teams to all the personal Perspectives, and then converging back via team comparisons to the group comparison Perspective. This models a collaborative knowledge-building process that combines divergent brainstorming and convergent negotiation.
Functionally considered, a KBE with Perspectives like WebGuide consists of two primary subsystems: a Perspectival data selection computation and a set of interface displays of the selected data. When a display is requested, the system must search the database to determine which content notes should be displayed to the particular user in the requested Perspective. For instance, if I request to view your Perspective, the system must select notes that are defined within your Perspective or within any Perspective from which yours inherits (recursively), except for notes that are private or that have been over-ridden. Various special displays can also be computed using this inheritance computation by treating discussions, negotiations, historical archives, etc. as pseudo-Perspectives that have special inheritance and exclusion rules. Once the Perspectival data computation has been returned, the content can be displayed in specialized interfaces that provide different kinds of functionality useful for further knowledge building.

**An Open Source Perspectives Server**

A specific task of the proposed project is to separate out the Perspective computation from WebGuide and structure it as a self-contained module with a well-defined application programming interface (API). This will form a Perspectives Server, a Java application that runs on the Web server along with the database system. It will be separate from the WebGuide client that will still run in a Web browser on the client’s computer. This separation of functions into a server and a client will have many advantages. It will speed the functioning of WebGuide because the intensive computation of Perspective content will be done on a central server that is faster than typical student computers. Also, calls to the database system will take place locally rather than across the Internet. In terms of system development, it will mean that developers can build systems that incorporate Perspectives without having to worry about the Perspective algorithms or the database calls. They will use an API that lets them request data that should be shown to a given user in a given Perspective. They can then just focus on how best to display this data in the interface.

The Perspectives Server will be a self-contained Java application. It will be released as open source with clear documentation on how to use it to get Perspective data for display. The data will be delivered as an XML text stream that can be used by any Web technology, such as HTML, Perl, or Java. The data will be human-readable, making it easy for programmers to see what data is being passed. Although it is anticipated that the Perspectives Server will generally be used as a black box, its open source availability will allow programmers to modify it if necessary, such as to
incorporate improvements to the XML DTD or in response to changes in Web technology. However, the Perspectives Server will be designed to make expansions of the database schema easy to incorporate without changes to the source code. This will allow new data structures corresponding to new multimedia data types.

The Perspectives Server will be a form of middle-ware, operating between the database and the client software. It will instantiate a three-tier, model-controller-view architecture that defines independent layers for the data schema or model, the data computation or control, and the interface display or view. The database management system can be any standard relational SQL system like mySQL or Oracle. The middle layer can be the Perspectives Server or a stripped down version that does not compute Perspectives. And the interface can be any kind of applet, Web page, or Web application that conforms to the API standard.

### A Component Architecture for KBEs

The release of the open source Perspectives Server will not only facilitate the rapid prototyping of Perspectives-based systems for this project to use in its study sites, it will also allow other researchers to incorporate computational Perspectives in their KBEs. We have already had requests for this from researchers in California and in Germany.

The PI of this proposal is involved in several efforts to promote collaboration among KBE researchers. Among these, he is the PI on a seed grant from the NSF-funded Center for Innovative Learning Technologies (CILT) to foster data interoperability among KBE systems. Work on this is currently producing a draft XML DTD (eXtensible Markup Language – Document Type Definition) to provide a common data format that KBE data can be imported to and exported from. Data in this format can be displayed using CSS and XSL. Tools we are now developing will allow such data from any KBE to be analyzed with standard measures, allowing for instance the volume and characteristics of discussion threads on different systems to be quantified and compared.

This DTD will also provide the format for data transfer between the Perspectives Server and interface clients. Clearly, interface clients will have to be designed to accept and make use of data received in this XML format. An important aspect of the proposed work will be to define a set of standards: the XML DTD for data interchange, the API for the Perspectives Server, and the ability of interface clients to call the Server and make use of the data.
The PI will be communicating with other KBE researchers nationally and internationally to solicit their concerns about these standards and to work toward a consensus and adoption of a set of such standards. These standards do not have to be formally approved by international standards bodies; informal agreement within a set of collaborating research groups is all that is needed for substantial practical benefits. Use of these standards will allow for rapid prototyping and customization of systems for various study sites both within the proposed project and by other researchers. It will also move us significantly toward a future in which KBE components from different research groups can be intermixed so that new systems can take advantage of functionality developed at different sites. Certainly, every effort will be made to incorporate related international standards, such as those for XML, XLINK, and metadata.

**Intelligent Hyper-Linking of Related Perspectives**

System development efforts in Year I will focus on release of the Perspectives Server and in Year II on prototypes for the study sites. In Year III, we will enhance the power gained by organizing notes into personal and group Perspectives by adding the functionality to locate the notes most closely related to a given note, such as a new idea just entered into my personal Perspective or an old note proposed for inclusion in the group Perspective. Research in KBEs like CSILE has shown that it is difficult to locate related ideas within a shared database of discussion notes (Hewitt et al., 1998; Hewitt & Teplovs, 1999).

We will use Latent Semantic Analysis (LSA) (Landauer & Dumais, 1997; Landauer et al., 1998) to analyze the semantic content of notes and to measure the semantic relatedness of pairs of notes. LSA is based upon a statistical analysis (singular value decomposition) of co-occurrences of terms in a large corpus of text. It determines the relatedness of words even if they did not occur together explicitly – hence the term “latent”. LSA incorporates some refinements that make its performance closer to that of humans than similar methods (see special issues of Discourse Analysis 1997 and Interactive Learning Environments 2000 on LSA assessment studies).

The PI and his graduate assistant are currently completing a four year project (sponsored by the McDonnell Foundation CSEP Program) that successfully uses LSA in a Web-based educational system tested in middle school classrooms (Kintsch et al., 2000; Stahl et al., 2000; Steinhart, 2000).

Automated linking of related notes will involve a fairly straight-forward application of LSA. It will be handled within the Perspectives Server, running on a computer with access to the necessary files for LSA. For each
major application or study site, a corpus of domain-related text will be subjected to LSA analysis to define a semantic space. Periodically (e.g., each night) the site’s shared database of notes will be folded into this corpus to redefine the space and to compute the vector for each note within this space. In real time, when linking is requested for a new note, the note’s vector can be quickly computed and a list of notes with the closest vectors in the semantic space can be produced without noticeable delay.

We will experiment with different interfaces to try alternative approaches to incorporating this functionality into KBEs. For instance, it can be left to users to ask for lists of notes related to a given note. Alternatively, an agent can automatically check to see if there are notes within a given closeness to certain notes: newly entered notes, notes proposed for negotiation, notes being read or edited, etc. The agent can then suggest that links be established from the given note to similar ones. The different interfaces can be tried out in our study sites.

2.3. Study Sites for Evaluating Knowledge Building in Workgroups

We will use local sites under our own control as alpha sites for testing our software, not only to eliminate bugs, but to try out different functionality and to refine the interface. More formal evaluation of the social impact of KBEs in workgroups will be conducted in corporate sites, primarily in the Boulder area. We have targeted two local situations (academic research and university learning) and two corporate sites (corporate training and industrial design).

Academic Research: the LD Center

The proposed work will take place within the Center for LifeLong Learning and Design (L3D), a research group within the Department of Computer Science at the University of Colorado. We will try out our prototypes in a variety of applications within L3D. Such self-application will give us first-hand experience with the requirements for the use of KBE software and with the practical problems of deployment and adoption. Increasingly, research at L3D involves participants from different disciplines and even virtual subgroups, like colleagues at other universities and other countries. When, e.g., a research project involves participants from Boulder, Colorado, and from Dortmund, Germany, a Web-based collaboration medium is essential, and means for defining and negotiating personal, subgroup, and whole-
group Perspectives seem particularly appropriate. Because users in this group are themselves software developers and researchers, they are particularly accepting of glitches and are reflective about design issues. This will provide a convenient and forgiving initial test site.

**University Learning: College Seminars**

Several members of L3D, including the PI, offer undergraduate classes and graduate seminars. These courses typically emphasize student on-line discussion and group projects. They often have a content focus on the Internet. We have used WebGuide and other Web-based tools in past courses and will continue to use them in the future. We are exploring courses that not only span multiple disciplines, but also span the oceans. Again, here, Perspectives for subgroups (interest-, content-, discipline-, or location-based) make sense. Collaborative classrooms will provide a secondary test site in which KBE functionality and prototypes can be tried out and knowledge-building activities can be monitored.

**Corporate Training: Athenaeum International DesignShops**

We foresee our primary study site for this project being a corporate training setting, in which knowledge building takes place under settings that may be advantageous for study. Athenaeum International (AI) is part of a distributed network of corporate training facilitators associated with MG Taylor. AI is located in Boulder and has established a good working relationship with the PI and with L3D. AI specializes in the design and manufacture of custom movable furniture for rapid deployment at corporate training events, as well as the facilitation of such events. They are interested in incorporating knowledge building software tools in their furniture to support the training process. (See attached letter of support from AI.)

A typical training event – or DesignShop™ – involves bringing together decision-makers from throughout a company to “reengineer” their corporation or re-think their high-level mission. This might involve a series of three-day workshops, or even an on-going sequence of quarterly gatherings. AI staff have noted a number of problems that they think could be addressed by innovative computer support: there is tremendous time pressure and everyone cannot express all their ideas and arguments; too much time is spent introducing materials; it is hard to retain important points and decisions; follow-through is tricky; documentation is labor-intensive. KBE support could include pre-workshop preparatory discussions, capturing of ideas that arise in the face-to-face meetings, organized documentation of
debates that took place, and follow-up discussion, analysis, negotiation, decision-making, or follow-through.

AI would provide a challenging study site for KBEs. DesignShops are high-stakes events involving people who need to make efficient use of their time. The groups here would be larger than the test groups at the university, and would involve more intense face-to-face interaction. A workshop series is of limited duration, so success could be assessed quickly and changes made prior to a subsequent trial. Project staff would have the aid of AI’s experienced group process facilitators to guide the design of the software support and of the deployment strategy, as well as to analyze the impact the software had on the social systems.

Industrial Design: Seagate Technical Design Centers

Another targeted corporate study site provides a rather different opportunity for investigating the use of KBEs. We have begun to investigate a particular work group within Seagate, a major hard disk manufacturer located near our university. Seagate is an established high-tech corporation. Through a series of acquisitions and mergers, it now consists of large design centers in Longmont (outside of Boulder, Colorado), Minneapolis (MN), Kansas City (IW), Redwood City (CA) and Singapore. Each of these design centers houses a few hundred employees working rather independently of the other centers.

We will focus on the effort of a Seagate Vice-President who is in charge of coordinating research on the problems of disk drive head tracking. In order to meet market demands that are projected as exceeding Moore’s law – requiring a doubling of storage density every year – the designers who work on head tracking must solve complex issues in physics and mechanical engineering. Unfortunately, engineers at different centers work almost independently of each other, duplicating research and designing products that overlap in functionality and specifications. As a result of their different histories, the centers have very different cultures of work, interaction, and outlook. For instance, one center prides itself in minimizing costs while another spares little cost to make what they consider a quality product. Designers from the different centers are accustomed to different engineering paradigms and find it hard to talk to each other.

The research objective here would likely be to use KBEs to structure communication and collaborative design among the distributed design groups. There is already a shared Lotus Notes database in which everyone can view the specifications and deadlines for each group’s product line.
However, there is currently no medium of communication among the groups (other than generic email) and no persistent textual discussion of the posted data. This makes it hard to share interpretations, work on establishing common understandings, or collaborate on building knowledge from the data. There is no support for Perspectives that would represent the conflicting cultures of the distributed groups and allow for negotiation of these differences.

Other Potential Study Sites

We will try to work with both AI and Seagate initially to explore their suitability to this project. We may end up focusing on one or the other site for practical or theoretical reasons. It is always hard to predict how field research will develop over years, particularly in today’s volatile marketplace, where key contacts change jobs and companies alter their strategies. For this reason, we have also established access to several other industrial sites similar to AI and Seagate: StorageTech is another major manufacturer of digital storage media; our contacts there are especially concerned with making their company a “learning organization” in which knowledge building is recognized to be an essential aspect of work. IBM has a major support center near Boulder, with a help-desk organization of 700 employees who must continually collaborate to build knowledge of the products they support. L3D has maintained a long-term relationship with two Japanese software companies, SRA and PFU, who are both interested in incorporating our ideas about lifelong learning and KBEs into their operations. These companies all have interesting settings where we would be welcome to deploy and observe our system prototypes if we have time during the proposed project.

Evaluation of Social Impact of KBEs at Study Sites

Evaluation will be conducted using converging methods to understand the complex, systemic issues around new technology deployment and use. Evaluation goals are two-fold:

- **Constructive:** We want to understand the environment with the objective of tailoring the design of the KBE technology to the study site, as well as constructively guide the deployment toward successful adoption.

- **Objective:** We want to objectively observe evolving use of the KBE technology; assess the nature of the IT impacts on coordination and collaboration; and evaluate the validity of the "collaborative knowledge
building" concept and applicability of the KBE theory; and refine the KBE theory based on results from observed practice.

Data collection will take the form of face-to-face in-depth interviews; phone and email-based "interviews" when subjects are at a distance (particularly in the case of Seagate); real-time non-participant observation in the workplace as well as via the KBE technology; surveys (particularly in the case of the AI training sessions, where there will be many more subjects); KBE database data collection; and document collection where appropriate.

In addition to coding and content analysis of field and interview notes, analytical approaches will also include structural analysis of discussion thread lengths and participation levels based on the KBE database data. We will perform discourse and content analysis of argumentative exchanges, and KBE-captured collaboration episodes.

Issues we will investigate include:

- What is the nature of collaborative knowledge building, and what activities comprise it?
- How does asynchronous support for articulation affect participation with respect to a variety of factors including time, location, and social status?
- How does computational support for discussion, argumentation, and clarification affect consensus-building as well as conflict?
- Do subjects understand and put into practice the concept of computational Perspectives? How do subjects interact with and manage multiple Perspectives (personal, subgroup, group, comparison)?
- Can "synergistic moments" be captured? What portion of these moments occur on-line in KBE environments?
- What role do facilitators (in the case of Athenaeum International) and management play in the use of KBE technologies?
- What are the particular hurdles that must be overcome for KBEs to be useful in these particular organizational environments?

Since the evaluation is a multiple-person effort, and because there will likely be different students participating at different times during the pre-and post-deployment stage, we will carefully organize observation efforts, and make an effort to systemize field note format to the best possible extent. We will
also conduct regular group data analysis meetings to coordinate the results of our efforts.

3. OBJECTIVES OF PROPOSED WORK

3.1. Objectives for Theory Development

- Investigate the phenomenon of the "synergistic moment" in interactions captured in the database.
- Investigate the utility and actual usage of computational Perspectives.
- Refine the model of collaborative knowledge-building activities based on project findings.
- Publish results of this project in a monograph.

3.2. Objectives for Technology Prototyping

- Define a standard for data interoperability among KBEs.
- Release open source import/export/display/analysis tools for KBE data interoperability.
- Release an open source Perspectives Server.
- Release example KBE interface components using different technologies.
- Develop an agent for intelligent hyper-linking of KBE notes.

3.3. Objectives for Deployment

- Deploy a KBE prototype in settings of academic research (e.g., L3D research group).
- Deploy a KBE prototype in settings of university learning (e.g., seminars at the university).
- Deploy a KBE prototype in settings of corporate training (e.g., AI DesignShops).
- Deploy a KBE prototype in settings of industrial design (e.g., Seagate design groups).
3.4. Objectives for Evaluation

- Conduct initial evaluation of the way work is presently conducted among selected groups to: collect baseline data; carefully select groups to deploy to; and formulate deployment strategies.
- Assess existing challenges for communication and collaboration within and across groups through interviews and possibly email-based interview-type surveys for those participants at a distance.
- Observe deployment of KBEs and revise deployment strategies as necessary.
- Perform on-going qualitative (observation, interviews, and document collection) and quantitative (KBE database activity) data collection.
- Analyze collected data to evaluate the impact of the KBEs in the workgroups.

3.5. Objectives for Dissemination

- Build a local, interdisciplinary community of students and faculty to conduct KBE research.
- Collaborate with at least 3 US research groups and with at least 3 international KBE research networks.
- Maintain an active website with the results of this project, including the open source products.
- Present the work of this project in the Group ’01 and CSCL ’01 international conferences (to be held at the University of Colorado), and report the findings of this project at 3 or more other international conferences.

4. PLAN OF PROPOSED WORK

4.1. Year I

Refine the model of collaborative knowledge-building activities through presentations to researchers and analysis of the "synergistic moment" in face-to-face collaboration.

Define a standard for data interoperability among KBEs using XML and XLINK for notes and relations among notes, including threaded discussions;
explore the adequacy of this standard using local prototypes; circulate the proposed standard among other KBE researchers; integrate the proposed standard with emerging data standards.

Define a standard for KBE interface components including JavaBeans, using XML for data interchange; explore the adequacy of this standard using local prototypes; circulate the proposed standard among other KBE researchers; integrate the proposed standard with other emerging data standards.

Define a standard for KBE database servers to provide XML data structures to interface components meeting the above standards.

Release an open source Perspectives Server that efficiently computes data visible in a requested Perspective in the standard XML format. Provide open source and documentation so that other researchers can use this Server for Perspectives-based KBEs and can propose improvements to the Server.

Deploy a KBE prototype in a setting of academic research such as a local research group. The KBE will be used to discuss and design standards and software. Learn from this deployment experience.

Begin initial observation of two primary test sites (AI and Seagate) to prepare for eventual KBE deployment.

Collaborate with American research groups interested in KBEs (e.g., Berkeley, SIU, Stanford, Georgia Tech, Hawaii, SRI) and international KBE research networks (e.g., existing CSCL research networks in Germany, England, Norway, Finland, Canada) to coordinate software component design around agreed upon standards.

Establish a website to publicize the results of this project, including the open source products, and to provide an information center for collaborations. Incorporate a KBE in the website to foster interactive knowledge building.

### 4.2. Year II

Release open source import/export/display/analysis tools for KBE data interoperability, using XML to define a DTD and using XSL and CSS to display the data. Document the standard and the tools on a website that makes them available to other researchers.

Release example KBE interface components in different technologies, including Java, Perl, and HTML. Make these available on a website with documentation and tutorials to help other researchers develop compatible components.
Deploy a KBE prototype in a setting of university learning such as a seminar offered by the PI, and observe use. The KBE will be used for students to develop, exchange, and negotiate reflections on shared readings and on collaborative writing projects.

Continue investigating the AI study site and assess the requirements for deployment of a KBE there.

Deploy a KBE prototype in a setting of corporate training such as a series of AI corporate training sessions. Commence post-deployment observation.

Present approach and intermediate findings of this project at the Group ’01 and CSCL ’01 international conferences. During Year II, these conferences will both be hosted at CU, with the PI serving as chair of the local arrangements committee at Group and chair of the program committee at CSCL.

4.3. Year III

Deploy a KBE prototype in a setting of industrial working such as a technical design group at Seagate, and observe and analyze use there. The AI deployment will also continue to be evaluated in Year III.

Develop an agent for intelligent hyper-linking of KBE notes using LSA technology to locate notes in a KBE database that are semantically related to a given note.

Continue on-going analyses of KBE use and examine findings with respect to the project's central hypotheses about KBEs. Publish a monograph on the project findings.

Report on this research at 3 or more international conferences.

5. EXPECTED IMPACT OF PROPOSED WORK

The proposed work – with its development of theory, defining of architectures and standards, prototyping of systems, deployment at several study sites, evaluation of utility, and dissemination to collaborators around the world – may seem overly ambitious for a half-time (6 months per year for the PI) project. This would certainly be the case if a lot of the groundwork had not already been laid and if there was not an expectation that this project will take place within a context of supporting work. The PI has already published
on KBE Theory and begun to sketch more in his seminars. Work is already well underway on separating the Perspectives Server out of WebGuide and defining an appropriate XML DTD for the data format – although these two projects have not yet been combined. Building a consensus on technical standards for KBEs has been initiated through the CILT grant and a popular workshop at CSCL ’99. It will continue at planned conferences, including Group ’01 and CSCL ’01 where the PI will be involved in planning. WebGuide and other L3D prototypes have been used and evaluated in middle school and college courses, including the PI’s recent seminars. Colleagues in L3D are developing systematic methodologies for analyzing and comparing discussions in KBEs.

The proposed project is conceived as the centerpiece for the PI’s work in the coming years. A large NSF ITR pre-proposal is currently pending that would support the collaboration efforts associated with this work. Another proposal to NSF’s CSS program would continue our OMOL project, orienting it particularly toward the deployment and adoption effort at the corporate study sites and at the analysis of the social impact of the KBE software in these settings, while a pending CISE Research Infrastructure proposal includes equipment and infrastructure for the local academic study sites. If funded, these other proposals would provide adequate support to fully achieve the vision proposed here. If they are not funded, the PI will seek other grants leveraged on the work proposed here.

The proposed work is designed to provide some of the basic conditions for the sorts of intensive collaboration that we believe are necessary to achieve the potential of KBE technology. In addition to providing technical conditions (e.g., Server, architecture, standards) and social conditions (local, national, international networks), the work will prepare specific examples of functionality (e.g., computational Perspectives, automated linking) and concrete analyses of application. These results will be disseminated through communication channels including specific related conferences.

Within the PI's home institutions the proposed work will be an important catalyst for building an interdisciplinary collaborative effort of undergraduates, graduate students, research staff, and faculty working on KBE theory, technology, and evaluation. The proposed work defines a coherent research agenda over several years that will crystallize a local research community.
References


Proposals for Research


Stahl, G. & Herrmann, T. (1999) Intertwining perspectives and negotiation, In: Proceedings of International Conference on Supporting Group Work (Group '99), Phoenix, AZ. Available at:

http://www.cs.colorado.edu/~gerry/publications/journals/c&e/.


Collaborative Research on Knowledge-Building Environments: Growing a National and International Research Community for Distance Learning Information Technology

Knowledge-Building Environments (KBEs) are software systems to support collaborative distributed learning. This is a complex research area that has made significant progress in the past decade but that will require substantial work by an international research community to achieve its potential in the next decade. Active research networks have been established in many countries, but there is no organized network of KBE researchers in the US to work collaboratively within this international community.

Coordinated multi-disciplinary work is needed at the levels of theory refinement, software design, and curriculum development. Many of the necessary enabling elements are becoming available now for progress in developing this new KBE information technology that will meet rapidly growing societal requirements: theories of learning that recognize the role of social context; technologies for building and combining software components; experiments in structuring effective distance collaborative learning; and networks of researchers in other countries. This project will build on these elements in the following areas:

- **Learning Theory:** synthesizing theoretical approaches into an analysis of social knowledge building, oriented toward the design of software to support collaborative distance learning;
- **Information Technology:** defining technical standards for the interoperability of KBE data,
- **knowledge-base servers, Web interface components, and agent widgets from different**
- **research prototypes;**
• IT Education: developing curriculum, course methods, and assessment measures for educating multi-disciplinary students in the theory, design, and educational use of KBEs;

• IT Workforce and International Collaboration: involving students and researchers in hands-on software design and the sharing of design ideas in face-to-face workshops and in Web-based discussion, and growing a network of researchers and students in the US to work with researchers abroad.

The project will expand over five years from one grantee (Colorado) and three subawards (Berkeley, Cornell, Southern Illinois) to five collaborative grantees and five subawards, directly supporting from 9 to 21 students each year and involving many more in courses and conference workshops. The project will create an active research community, involving educators and technologists together, enhancing the utility, scope, and depth of IT support for knowledge-building activities. This will create a workforce capable of turning the potential of distance learning into a classroom and workplace reality, using the theories, technologies, and methodologies developed in this project through international collaboration.

Problem Statement

“Long-distance learning is skyrocketing” according to an Associated Press article (December 19, 1999). Already in 1997/98, 60% of colleges offered Internet classes, with 54,000 different courses enrolling 1,600,000 students. Yet, the technology and methodology for designing Internet courses is still very poorly understood. Most teachers simply put traditional course materials on the Web, ignoring the potential of computational support. At best, they use generic communication technologies (like First Class, NetMeeting, Lotus Notes/Domino) that were not designed to support learning, or they use systems that administer and deliver traditional materials (like WebCT or LearningSpace) but do not go beyond this (Cameron et al., 1999).

The educational research community of the past decade has established a consensus that traditional lecture-based and teacher-centered approaches do not by themselves produce the most effective learning. Students should be actively involved in constructing their own understanding within collaborative social contexts. Students in a course should function as a community of learners, community of practice, or knowledge-building community (Brown & Campione, 1994; Brown & Duguid, 1991; Lave,
Proposals for Research

1991; Scardamalia & Bereiter, 1996). Active student projects that provide authentic motivation can form the core of a problem-based learning (PBL) approach (Barrows, 1994). Computer-based tools should be designed to support the collaborative knowledge-building process. Although there is broad agreement that methodologies and tools are needed for computer support of collaborative learning (CSCL), these have yet to be developed.

Important initial steps have been taken to formulate theories, try out prototype systems, develop pedagogical methodologies, and experiment with innovative courses (Hoadley & Roschelle, 1999). These steps have provided enough experience to demonstrate how much is left to do and to indicate a path for further research. There is an assortment of theoretical approaches that seem vaguely complementary, but no synthesis that provides a coherent framework for designing courses and knowledge-building environments (KBEs). One software system – CSILE (Scardamalia & Bereiter, 1996), now Knowledge Forum – has been under development for over a decade and has been widely fielded and assessed. A number of other prototypes are being designed and investigated to explore alternative functionalities: CoWeb (Guzdial et al., 1999), WEBGUIDE (Stahl, 1999b), WISE (Cuthbert, 1999), CoVis (Pea, 1993), FLE (Muukkonen et al., 1999). These attempts to support collaboration repeatedly run into the same technical and social problems: low participation levels, shallow discussions, divergence of ideas, little building of deep knowledge structures (dePaula, 1998; Guzdial et al., 1999; Guzdial & Turns, forthcoming; Hewitt et al., 1998; Hewitt & Teplovs, 1999; Stahl, 1999a, 1999b, 1999c). The overcoming of these barriers to collaborative learning remains an open research issue.

Experience indicates that the design of the KBE “killer app” is too complex for any one research group. The theoretical, technical, and pedagogical issues are deeply intertwined and each still requires basic research. A high-functionality system is needed, unlike the self-contained functions of email, the Web, or e-commerce. An international research community is emerging to address this challenge, with energetic research networks and international virtual universities in a number of countries. Unfortunately, there is no coordinated effort within the United States which can relate to these networks abroad. We need to develop a multi-disciplinary community which can understand and advance the theory, technology, and pedagogy; can disseminate that understanding in carefully conceived courses; and can interact in the international community.
Project Goals

Learning Theory: To synthesize and adapt current theories of computer supported collaborative learning to define a conceptual framework for the design of knowledge-building environments.

Information Technology: To propose, negotiate, and promote interoperability standards for data and components of knowledge-building environments.

IT Education: To develop and test content and methodology for multi-disciplinary, problem-based courses on information technology for distance learning.

IT Workforce and International Collaboration: To build a US network of established researchers and new students in the field of computer supported collaborative learning that can collaborate with networks in other counties on information technology for distance learning.

Research Issues

Learning Theory: How can current theories be synthesized into a coherent view of knowledge-building processes and how can this guide the design of software?

Information Technology: How can standards be defined for interoperability of KBE data, knowledge-base servers, Web interface components, and agent widgets to promote exploration without restricting software design options?

IT Education: How can problem-based learning be adapted to distance learning? What software can support this? What constitutes an effective curriculum (problem case-base) for coverage and depth concerning information technology for distance learning?

IT Workforce and International Collaboration: How can a productive network of US researchers be established, grown, and sustained so they can collaborate with distance learning research networks in other countries?

Project Objectives

Learning Theory: To produce a series of white papers that are discussed by the project community and then published.
Information Technology: To establish a set of interoperability standards, examples, and tools.

IT Education: To develop and test a sequence of courses on the technology and pedagogy of distance learning.

IT Workforce and International Collaboration: To organize periodic workshops for project members, students, and international collaborators and to provide Web-based media for project reports and discussions between workshops.

## Theoretical Framework

This project focuses on a particular approach to CSCL – namely the Web-based support of collaborative knowledge building (KBE) – and a particular approach to instructional design – namely problem-based learning (PBL). While face-to-face PBL is an established method, the design of distributed PBL using KBEs is very much a current research topic (Cameron et al., 1999). To provide a framework for the design of KBEs to support distributed PBL, the project will synthesize and refine a set of currently accepted theoretical approaches from the perspective of guiding software design.

The diagram below (from Stahl, 2000) provides a starting point for this, combining aspects of activity theory, situated learning, hermeneutic philosophy, and distributed cognition theory.

The idea of this diagram is that the knowledge-building process can proceed through many different phases. A KBE can be designed to support a number
of these phases with different functionality. A similar approach is being developed in Finland (Muukkonen et al., 1999).

## Technology Approach

Many ideas of KBE functionality have been tried out and a number of promising new features and approaches have been proposed. The problem is how to combine various sets of features into a technically and pedagogically coherent system. In order to allow functions from different prototypes to be recombined to explore new system configurations, we need to achieve interoperability of data, servers, components, and widgets.

The PI has already begun to define an XML standard for interchange of threaded discussion data, which forms the core of many KBEs. Data from four different prototype systems used at Colorado have been exported to the XML standard, where they can be displayed in XSL and analyzed by simple text manipulation tools. Work has already begun on separating a perspectives server out from the PI’s WEBGUIDE KBE (Stahl & Thomas, 1999), so that Web client interfaces developed using HTML, Perl, or Java can easily access a shared knowledge-base without worrying about the database or perspectives computation internals. Java beans technology provides a technical foundation for programming components and widgets that can be mixed and matched in alternative systems. So the technology for interoperability seems within reach.

The problem is to agree on standards within the KBE community. The goal is for someone to be able to combine, for instance, a knowledge-base server from Colorado’s WEBGUIDE, a discussion interface from Toronto’s CSILE, domain scaffolding from Berkeley’s WISE, and a design module from Helsinki’s FLE with some innovative notification agent widget. Then they can assemble a system to test the effectiveness of their new agent widget (McLean, 1999) without having to build a whole system from scratch. The data from their experiment can then be exported to XML and analyzed with existing tools to compare the results with those of other systems.

The definition of interoperability standards requires international collaboration. Although many researchers are informally converging toward a common set of technologies (SQL backends, Java servers, Web-based clients, threaded discussion), the KBE software field is still very immature. It will be important to devise standards that foster experimentation rather than restrictions that limit design options. That is a tricky research issue.
Curriculum Development

The starting point for curriculum development in this project is provided by the work of the Problem Based Learning Institute (Cameron et al., 1999) and the Canadian CollabU (Breuleux et al., 1999). The PBLI has tried to support distributed PBL with commercial communication tools, and has identified specific needs for customized KBE software. CollabU has begun to experiment with a course on learning technology taught at five different universities, with students divided into cross-campus projects. The PI has also conducted two multi-disciplinary seminars using KBE prototypes for class discussion: one on the theory of KBEs and one project-based course of KBE research with students at Colorado and Dortmund.

This project will begin by working with PBLI and CollabU, participating in their experiments and offering our own multi-campus courses. Our courses will focus on the wicked problems of KBE software design and will use various KBE prototypes.

During the grant period, we will develop both a curriculum and an instructional methodology for courses on information technology for learning. The methodology will define an approach to distributed learning design, incorporating and adapting techniques that have proven successful in face-to-face PBL. The use of appropriate technologies will be described. The methodology will emerge from our experimental courses. Course content will cover theory, pedagogy, and technology. It will be aimed at a multi-disciplinary undergraduate and graduate audience, as well as at classroom teachers, distance education instructors, and workplace trainers.

Community Building

The PIs of this proposal will be hosting the next CSCL conference (December 2001) and the next GROUP conference (October 2001) at the University of Colorado. Project participants will also be active in the European CSCL (December 2000) at Masstricht in the Netherlands, as well as meetings of ICLS (International Conference of the Learning Sciences), CILT, AERA, CHI, Cognitive Science, WebNet, CSCW, and other important international meetings of computer science and education researchers.

At CSCL ’99, the PI (with Marlene Scardamalia and Timothy Koschmann) planned and conducted a successful workshop with over 60 participants from the US and abroad on “Collaborating on the Design and Assessment of
Knowledge Building Environments in the 2000’s”. Many of the ideas and prospective participants of this project were involved in that workshop, which itself grew out of an earlier working group at CILT ’99 (the NSF-supported Center for Innovative Learning Technologies). This proposal is a product of collaboration funded by a CILT seed grant intended to stimulate collaboration among KBE researchers.

International conferences provide a convenient venue for national and international collaborators to meet face-to-face as a supplement to Internet-mediated communications. This project will organize four meetings per year for project participants and collaborators to get together. Two of these will be day-long organized conference workshops where people will exchange and discuss their project work results. The other two will be informal SIGs where people can socialize and exchange ideas one-on-one. Some of the meetings will be in Europe; some will take place at research labs. The project will support some costs of student participants to attend these meetings and international conferences.

The building of a national and international collaboratory is an explicit aim of this project. In addition to involving people as project participants (collaborating PIs, subaward recipients, student researchers) and as students in project courses, the on-going work of the project will be publicized widely. Articles in conference proceedings and journals will be important, with shared focus and special tracks or special issues arising naturally from the project identity. In addition, a project newsletter will be circulated by email and Web, and discussion forums on project topics will be supported by the KBEs that emerge from the project.

**National Collaboration**

In years 2 through 5, from 2 to 4 other universities will be added to this project as collaborative institutions. Each will submit a collaborative research proposal with biographical information about their PI and a budget to support a graduate student and an undergraduate student to work on this project. Each collaborative budget will run about $120,000 per year. Collaborating institutions will be selected as part of the work of the project, with the following institutions tentatively interested at this time (more details in full proposal):
International Collaboration

There are now active KBE research networks in the following countries who have expressed strong interest in collaborating with this project (more details in full proposal):

- **Canada**
  - OISE Toronto
  - Marlene Scardamalia & Robert McLean

- **Finland**
  - Helsinki & Turku
  - Kai Hakkarainen

- **Norway**
  - Oslo & Bergen
  - Anders Morch

- **Germany**
  - GMD & Dortmund
  - Wolfgang Prinz and Thomas Herrmann

- **Canada**
  - TeleLearning NCE
  - Alain Breuleux and Tom Calvert

- **Mexico**
  - Monterrey Virtual U
  - Jose Rafael Lopez Islas

- **United Kingdom**
  - Open University
  - Simon Buckingham Shum

**Anticipated Impact**

This project will establish a new research focus on KBE design and a national network of established researchers and new students that will significantly contribute to an existing international collaboration exploring this field of information technology. It will produce enabling theories, technologies, and pedagogies to support the efforts of this new workforce to move KBEs from research prototypes to robust IT systems that can fulfill growing societal requirements. Effective KBEs will provide a new paradigm of collaborative knowledge management, exploiting the online availability of information with more powerful means than are currently available.
Papers from CSCL ’99 are available at the conference website. Papers by the PI are available at: www.cs.colorado.edu/~gerry/publications/


Phoenix, AZ. Available at http://www.cs.colorado.edu/~gerry/publications/conferences/1999/group99/.
IT Support for Knowledge Building in Workgroups

This project develops theories and technologies for collaborative knowledge building in workgroups. It:

• proposes a model of the activities through which knowledge is co-constructed within groups;
• uses the model as a framework for the design of IT support for knowledge building;
• produces a technical infrastructure for prototyping software knowledge-building environments;
• deploys customized prototypes in organizational settings for observation;
• assesses the deployment, adoption, and use of prototypes in actual workgroups;
• feeds findings back into the theory and technology.

Collaboration as Knowledge Building. Knowledge building is conceived here as an extension of the notions of organizational learning, organizational memory, and knowledge management. In particular, the focus is on the creation and capture of knowledge generated via human collaboration. An assumption of this work is that knowledge created by groups of people is of a different quality than when individuals work alone. It is popularly believed that collaboration supports the creation of better information; although this is often the case we know that it is not always true. However, we do believe that collaboration can generate information, beliefs, and knowledge that are public and shared – qualities often absent in the information that organizational memory systems try to embody. We propose that collaborative knowledge building consists of many forms of activity which require and promote such public sharing of meanings.

A Framework for Supporting Knowledge Building. The goal of this project is to develop, refine, and harness a set of collaboration technologies...
that can support different kinds of working and collaboration environments for the purpose of knowledge building. Our theory – based on cognitive, educational, philosophic, and social analyses – is designed to provide a framework for guiding this software development. It argues that computational systems can be designed to: (i) provide powerful external memories that extend human cognition, (ii) provide communication media that support group collaboration, and (iii) provide enabling technologies for collaborative learning. The theory's model suggests a variety of collaborative knowledge-building activities that can be supported with appropriate software functionality. Software to support individual activities has already been developed by members of the project group, by other research groups, and by commercial firms. However, there has been little effort to combine the various support functions systematically and to tailor them to collaboration in specific social settings.

**Knowledge-Building Environments.** This project will develop software components and architectures specifically to support collaborative knowledge-building activities, to facilitate the custom combination of components for particular settings, and to model the interpersonal relationships of the collaborators using the software. Components corresponding to the various activities will be designed to conform to standards for interoperability of data and function. A three-tier model-view-controller architecture will be established based on open system conventions so that an assortment of interface components using various Web technologies can display, manipulate, and modify the same underlying data structures. The data itself will be structured within a network of computational perspectives that represent the individuals and groups who have constructed the knowledge in the system. The software components can be easily modified and combined to allow for rapid prototyping customized to specific deployment settings.

**Study Sites and Evaluation.** Two sites with different organizational structure will be the targets of this work. We intend to evaluate and actively support IT deployment and adoption, to evaluate the usefulness of the IT in knowledge building work practices, and to better understand the processes and products of knowledge-building activities. The primary study site will be a company that conducts corporate training workshops. Here, experienced training facilitators will use project prototypes to extend the knowledge building that takes place in high-stakes corporate reengineering group decision-making activities. Another industrial setting is a hard disk manufacturer whose design teams are distributed in the world as a result of corporate acquisitions. Today, engineers at the different sites rarely work together, resulting in duplication of effort and incompatibility of results. At
both sites, we will investigate how knowledge-building environments with perspectives can mediate cultural differences among subgroups and move from sharing data to sharing understanding and knowledge. Real-world issues that affect these groups – such as individual and organizational incentives, corporate and cultural differences, time and money constraints, personal style differences and interpersonal relationships – will set the IT design and deployment agenda for this project.

Project Description

This project is guided by a theory of collaborative Knowledge-Building Environments (KBEs) that we are developing. This theory proposes the following principles:

• Collaborative knowledge building is a particular view of group learning that focuses on a range of activities that take place within communities, as opposed to focusing on learning as the transmission of bits of information to individual learners.

• Collaborative knowledge building takes place largely through the interaction among people with different understandings from multiple personal and group perspectives.

• Such knowledge building within groups can be helped by appropriately designed information technology (IT) that supports various knowledge-building activities and supports interaction among alternative perspectives.

The form of IT that we are interested in – collaborative Knowledge-Building Environments – represents a distinctive approach that overlaps related work in Computer-Supported Collaborative Learning (CSCL) and Computer-Supported Cooperative Work (CSCW). IT support for learning is traditionally oriented toward the transmission of information to individual students. Even where it is based on a view of student construction of knowledge, as with Intelligent Tutoring Systems (ITS) for algebra or physics, the goal is measured by testing the incorporation of pre-defined content or methods into the individual’s understanding (Wenger, 1987). A more student-centered, constructivist approach is taken by Interactive Learning Environments (ILE), which might, for instance, allow students to create ecologies in SimLife to learn biology, or programs in Turtle Logo to explore math concepts (Papert, 1980). In contrast, a KBE primarily supports the group process and leaves matters of content up to the participants (which may include a teacher who raises particular content issues and helps
maintain focus). In this way, it applies CSCW approaches to CSCL. A review of CSCW technology for groups (Kraemer & Pinsonneault, 1990) distinguishes group communication support systems (GCSSs) from decision support (GDSSs). GCSSs are specific communication media like email and video-conferencing. In providing computational tools for group decision making, GDSSs tend to support isolated, focused activities that integrate products of individual work. In contrast, a KBE aims to support a broad spectrum of knowledge-building activities – both individual and group – in a more seamless fashion. It supports the construction of areas of knowledge through group inquiry over extended periods of time. It also supports the interplay of individual and group more comprehensively, through integrated mechanisms of "computational perspectives" and negotiation that treat the group as more than just the sum of the individuals.

Assessments of CSCL and CSCW systems have defined a number of key issues for evaluating the problems and successes of such systems. For instance, in simple threaded discussion forums common problems include: short threads (a tendency for discussions to die quickly), low participation (lack of motivation to participate), few cross-references (little convergence of ideas), and superficial content (minimal depth of investigation) (dePaula, 1998; Guzdial & Turns, 2000; Hewitt & Teplovs, 1999). On the other hand, GDSSs and GCSSs attempt to decrease communication barriers within the group, while increasing task-oriented focus, depth of analysis, and decision quality (Connolly, 1997; Kraemer & Pinsonneault, 1990). Social informatics studies have raised additional issues of software deployment and adoption in addition to questions of usability and utility (Kling, 1999). These are some of the dimensions along which KBEs must be assessed within realistic learning and working social contexts.

To date, the PI and his colleagues have begun to develop KBE theory in conjunction with Web-based KBE prototypes that support many of the activities described in the theory and that have been tested informally in collaborative learning classrooms. In particular, computational support for personal and group perspectives has been developed and tried out. Support for computational perspectives was explored in the PI’s dissertation (Stahl, 1993a) and has since been refined and adapted to the Web (Stahl, 1999a). This work has been described in relevant CSCL and CSCW conferences (see references by Stahl). The proposed project will build on existing concepts and prototypes, extending them substantially by: implementing a technical infrastructure to support data interoperability, integration of functionalities, and rapid prototyping; deploying customized KBEs in specific study sites; observing the social impacts of these IT systems in the work settings; revising the theory based on empirical findings; and fostering
a community of researchers working on IT support for knowledge building in workgroups.

1. PREPARATION FOR PROPOSED WORK UNDER PRIOR NSF SUPPORT

1.1. Organizational Memory and Organizational Learning (CSS)

“Conceptual Frameworks and Computational Support for Organizational Memories and Organizational Learning (OMOL),” PIs: Gerhard Fischer, Gerry Stahl, Jonathan Ostwald, September 1997 – August 2000, $725,000, from NSF CSS Program #IRR-9711951.

This grant prepared much of the background for the proposed work. The OMOL project started from a model of computer support for organizations as Domain-Oriented Design Environments (DODEs) in which both domain knowledge and local knowledge are stored in the form of artifact designs and associated design rationale (Fischer, 1994). This CSCW model evolved into one of Collaborative Information Environments (CIEs), that emphasized the interactive, asynchronous, persistent discussion of concepts and issues within an organization (Stahl, 2000a). Gradually, interest in organizational learning aspects led to involvement in CSCL and the model of collaborative Knowledge-Building Environments (KBES) (Fischer et al., 1999). A number of software prototypes were developed to explore the use of the Web as a communication and collaboration medium. Of these, the most important for the proposed work are the following:

• DynaClass: A discussion forum for use in college courses. It features ties to DynaGloss and Sources as well as email notification and specialized displays (Ostwald, 1999).

• WebGuide: Differs from DynaClass in providing more control over rearrangement of notes; features computational Perspectives.

• DynaGloss: A system for defining technical terms and keywords and for debating the definitions and reviewing the history of debate; linked to DynaClass and Sources in that each term shows all the locations in these other systems where the term is explicitly referenced.

• Sources: A system for annotating bibliographical entries; uses terms from DynaGloss as keywords.
• InfoMap: An interface component for creating a graphical display of linked notes like a threaded discussion; providing convenient drag-and-drop functionality.

Work on this grant led to the focus on KBEs as models of computer support for organizational memory and organizational learning. In particular, it provided a number of different systems, each with useful functionality, and brought home the need to define component standards so the functionalities can be combined more flexibly. As we tested and deployed these systems, we confronted serious issues of adoption and focused our concerns increasingly on socio-technical and social informatics (Kling, 1999) issues: motivation, media competition, critical mass, social practices, seeding, management, re-seeding, convergence of ideas, peer-to-peer collaboration, deployment strategies. These issues led to a new research agenda (Stahl, 1999b) and this proposal.

1.2. WebGuide and Environmental Perspectives (NOAA)


This grant funded the initial implementation of WebGuide as an integrated Java applet KBE supporting personal and group Perspectives. It was a joint effort between the PI, a middle school teacher, and a research group at the National Oceanographic and Atmospheric Administration (NOAA) labs in Boulder. The NOAA group is experienced in developing educational Web sites for schools. The teacher taught an environmental science class in which he wanted to spend the year having his students interview various adults and construct a set of contrasting perspectives (conservationist, regulatory, business, community) on a particular local environmental issue that the students had previously been involved in. WebGuide was used by the students to collect notes on their interviews and to formulate personal and team perspectives on the issue. Results of this software trial were analyzed and presented at conferences (Stahl, 1999a; 1999b; 1999c; Stahl & Hermann, 1999). These findings led to a number of revisions of WebGuide, including the separation of the Perspectives mechanism from the Web interface, and recognition of the need for software architectures, standards, and components to support flexible rapid prototyping of KBEs.
1.3. Collaboration in KBEs (CILT)

“Interoperability Among Knowledge-Building Environments,” PI: Gerry Stahl, September 1999 – August 2000, $9,124.21, from NSF-funded Center for Innovative Learning Technology (CILT), Subcontract #17-000359 under NSF grant #EIA-9720384.

This is a current seed grant whose purpose is to stimulate collaboration among KBE research groups. Part of the intention of the grant was to prepare a proposal for fuller funding, such as the present proposal and its currently pending complementary NSF ITR proposals for “ITR/IM: Perspectives on Knowledge-Building Environments” and “ITR/EWF: Collaborative Research on Knowledge-Building Environments: Growing a National and International Research Community for Distance Learning Information Technology.” This grant has already resulted in a semester-long student project involving three graduate and three undergraduate students (one collaborating virtually from Germany) creating an XML DTD that defines a data format for data imported from several different KBE prototypes and displayed in a Web browser using XSL. The grant supported a workshop entitled “Collaborating on the Design and Assessment of KBEs in the 2000's” at CSCL '99 at Stanford. This workshop attracted over 60 participants and was preceded by an on-line discussion of 28 submitted position papers. This grant has led to the emphasis on collaboration among KBE research groups and the need to put into place some of the technical and social conditions for such collaboration (Stahl, 1999a), as proposed here.

2. OVERVIEW OF PROPOSED WORK

2.1. The Research Team

The Center for LifeLong Learning and Design (L3D) is an interdisciplinary research center at the University of Colorado at Boulder within both the Department of Computer Science and the Institute of Cognitive Science. It conducts teaching and research on the use of information technology for learning in school, workplace, and the community. It holds weekly colloquia on the social implications of IT and related themes. All members of the proposed project team are on the L3D staff and are currently working together on the NSF-CSS-sponsored project.

Gerry Stahl: Research Professor in Computer Science and Cognitive Science with doctorates in philosophy and computer science. Studied philosophy and
social theory at: MIT (with Dreyfus, Chomsky), Northwestern University (when it was the American center for continental philosophy), Heidelberg (with Gadamer), and Frankfurt (with students of Adorno, Horkheimer, Habermas). Studied computer science at MIT (with Minsky), Colorado (with Fischer, McCall), and on-the-job. Teaches seminars on theory of collaborative knowledge building and conducts research on IT support for learning and design.

Gerhard Fischer: Full Professor and Director of the L3D research center since its founding in 1994. Successfully completed many research projects with colleagues and graduate students, involving collaborations with industrial partners such as NYNEX, US West, PFU, SRA, IBM. Teaches courses on lifelong learning, organizational learning, design, and the Web.

Leysia Palen: Research Professor. Studied analysis of IT in social systems using ethnographic methods while working at UCSD (Hutchins) and UCI (Grudin), as well as at Boeing, Microsoft, and Xerox PARC in student and professional capacities. Conducted extended field studies of IT adoption at Sun and Microsoft (Palen, 1999). Currently conducting an NSF CSS project on the adoption of groupware calendaring systems, and examinations of time and information management practices more generally.


Rogerio dePaula: MS thesis on methodology for analyzing social adoption of on-line discussion forums (dePaula, 1998). Currently assessing on-line review process of JIME with Sumner. Also currently GRA on McDonnell Foundation CSEP grant with Stahl, Fischer, Kintsch, and Landauer, ending this year, assessing educational software in middle school.

Leo Burd: MS thesis on learning in an activity theory framework, and community work on social adoption of IT in poor communities of Brazil. Currently GRA on our NSF CSS OMOL project, ending this summer.

Undergraduates: Two students from L3D’s Undergraduate Research Apprenticeship Program will apprentice.
2.2. Theory of Collaborative Knowledge-Building Environments

Collaborative Knowledge Building

Information Technology (IT) is a broad field that can be conceptualized in various ways. Traditionally, the computer was thought of as a medium for storing and delivering data, that can then be used by people in their work. More recently, the computer (especially with the Web) has become a medium of communication, through which people share information and knowledge. This communication can take a variety of forms. In simple forms of e-commerce or on-line voting, people submit their decisions about a fixed list of choices. In chat and most email, people exchange greetings and opinions, generally without changing those opinions. Many systems in recent years have tried to support a particular form of communication or social interaction like brainstorming or decision-making – often with very positive results (Connolly, 1997; Vogel et al., 1987). We are interested in a distinct but broader process of communication which we term collaborative knowledge building. Here, groups of people construct new knowledge through interaction of their ideas and perspectives, usually eventually preserved in documents or other artifacts.

Our theory of collaborative knowledge building (Fischer et al., 1999; 1996; 1993/1998; Stahl, 1975; 1993a; 1993b; 1999c; 2000a; 2000b; Stahl & Herrmann, 1999) proposes a concept we call the synergistic moment; we intend to investigate the validity of this concept in the proposed project. The synergistic moment is the critical point during collaboration in which a group constructs meaning that transcends what any participant may have “in mind.” The shared understanding that is generated in this process is a subtle phenomenon: It does not mean that everyone is in complete agreement or even that each individual has the same internal cognitive representations of what is discussed. Rather, it means that a certain group view has been expressed. The unit of analysis for describing this is the group, and is manifested in the group's discourse. Individuals may agree to disagree with the group understanding, and careful investigation may reveal that individual understandings differ from the group's view (Hatano & Inagaki, 1991). The intersubjective "sharing" is not a correspondence or overlapping of individuals' mental content, but a coordination or interaction of their participation in joint socio-cultural activity (Matusov, 1996). The synergistic moment is an emergent property of the group dialog as a cacophony of voices (Bakhtin, 1986). It could easily pass unnoticed as a magical fount of creativity; to more deeply understand it likely requires "thick description" (Geertz, 1973) and detailed interaction/discourse analysis (Jordan & Henderson, 1995), and therefore presupposes that the interaction was
captured in some medium. Fortunately, the literature on CSCL contains a number of incisive analyses (Roschelle, 1998) of the synergistic moment, although they do not highlight it as such.

The synergistic moment is a result of perspective-sharing (Boland & Tenkasi, 1995), but at the group rather than the individual level. It overcomes the problem pointed out by Feltovich et al. (Feltovich et al., 1996), that any one perspective may limit the ability to comprehend creatively the complexity of a topic under discussion. What typically happens is that one person makes a statement from her personal perspective; someone else interprets that statement from his own perspective and responds accordingly; others continue this process so that the discourse consists implicitly of reinterpretations from various perspectives. The drive to establish intersubjectivity and shared knowledge is powered by socio-cognitive conflict and contention among perspectives according to studies by Piaget and his followers (Perret-Clermont & Schubauer-Leoni, 1981). The dialog proceeds through sequential turn-taking and attempts to repair “misunderstandings” as understood from particular perspectives and reinterpreted from others. Thanks to the human drive to impose coherent social meaning structures (Geertz, 1973), a synergistic group understanding emerges. This shared understanding can play a central role in the further activity of the group and can be more or less adopted by individuals into their personal perspectives. Although the synergistic moment seems to the participants to emerge spontaneously, it can be understood as the result of many identifiable knowledge-building activities, as represented in our model (below).

**Perspectives in Knowledge Building**

According to hermeneutics – the philosophy of interpretation – human understanding is fundamentally perspectival. We construct knowledge from our situated perspective in the world: our historical position, cultural tools, and practical interests (Gadamer, 1960/1988; Heidegger, 1927/1996; Stahl, 1975). Computational support for knowledge building can represent our interpretive perspectives with computational Perspectives (Boland & Tenkasi, 1995; Nygaard & Sørgaard, 1987; Winograd & Flores, 1986). (In this proposal, Perspective—with-a-capital-P will refer to the proposed computational mechanism that mirrors human interpretive perspectives—with-a-lower-case-p.) In this sense, Knowledge-Building Environments (KBEs) with computational Perspectives are designed to support the essential structure of collaboration. A key hypothesis of the proposed work is that KBEs benefit from an approach that represents the perspectival nature
of collaboration. A goal of the project is to facilitate the incorporation of a computational Perspectives mechanism in KBEs – both in our own prototypes and in the work of other KBE research groups around the world.

Computational Perspectives have been explored by the PI in a number of software prototypes, in his dissertation system, and in his theoretical publications (Stahl, 1993a; 1993b; 1995; 1998; Stahl & Herrmann, 1999; Stahl et al., 1995). In a single-user system, computational Perspectives may correspond to different domains or professional viewpoints on a design problem, such as electrical, plumbing, structural, and heating concerns in architecture (Fischer et al., 1993; 1993/1998). In a KBE to support collaboration, computational Perspectives typically provide personal or group workspaces for the development of different sets of ideas. In this way, they can model the relationships among the various personal and group interpretive perspectives at work in the construction of collaborative knowledge.

We hypothesize that computational Perspectives can support the synergistic moment in collaborative knowledge building by providing the necessary contact among different personal Perspectives, allowing them to interact, and then locating the results in a group Perspective. By situating the traditionally ephemeral synergistic moment within an explicit structure of computational Perspectives and by doing so in a persistent way, a KBE provides new opportunities for group self-reflection.

An important complement to Perspectives is negotiation. Negotiation is a process through which divergent personal perspectives converge on a collaborative shared understanding. When Perspectives and negotiation are effectively “intertwined” in a KBE, they compensate for each other’s potential problems: Negotiation converges ideas so that everyone can benefit from the ideas of other perspectives, while personal Perspectives allow people to work on their own views while potentially time-consuming negotiations are underway (Stahl & Herrmann, 1999).

For instance, when WebGuide – a KBE with computational Perspectives implemented by the PI – was used in a middle school environmental science classroom, students each had their own personal Perspective in which to develop their own responses to questions posed by the teacher. The teacher’s questions to the whole class were posed in the class’ group Perspective. From there they were automatically inherited into the team Perspectives. The content of the team Perspectives was, in turn, inherited into the personal Perspectives of team members. Gradually, students migrated their ideas to team Perspectives that represented either conservationist, governmental, corporate, or citizen perspectives on the
ecological controversy – depending on which perspective team the student was part of. Then they could work with the ideas of their team-mates and negotiate their team position. In the end, the different teams negotiated to spell out agreements and disagreements (Stahl, 1999c).

The analysis of the synergistic moment suggests that negotiation need not take the explicit, rationalist forms typical of GDSSs, such as voting. Group results may emerge naturally out of the intertwining of Perspectives in group discussion. A challenge of the proposed work will be to develop software support for capturing such results and migrating them un-intrusively to group Perspectives.

**The Potential of IT Support for Knowledge Building**

IT support has the potential of transforming the activities underlying the synergistic moment. For one thing, it would make those activities publicly accessible. The group could then reflect upon the emergence of its shared understanding by looking over the persistent record of its dialog. Such reflection might prove especially useful in contentious situations or for newcomers who were not part of the original dialog and are motivated to re-open the issue – as illustrated by Matusov (1996). Furthermore, computer support of perspectives could make explicit the interplay of different personal Perspectives and the migration of ideas and their interpretations between personal and group Perspectives. Ironically, perhaps, the “asynchronous” medium of the Web would allow group members to interact simultaneously – without waiting for sequential turns – thereby overcoming what Peters (1998) characterizes as “the hardest argument against democracy: the ability of only one person to speak and be heard at a time” (p. 261). Of course, as we have already discovered with the Web in general, the increased flood of ideas raises complex information management issues. We do not yet understand the full social impact of the envisioned KBEs – we will only know how they are used once they have been implemented, deployed in naturalistic settings, and observed.

Theories of human cognitive development emphasize the important role of external memories to extend short-term and long-term human memory (Donald, 1991; Norman, 1993). They also stress that individual cognition is a social product, highly mediated by social symbol systems, cultural artifacts, processes of structuration, and group collaboration (Bourdieu, 1972/1995; Geertz, 1973; Giddens, 1984; Hutchins, 1996; Vygotsky, 1930/1978). This suggests that computer support for collaboration has the potential to significantly advance the power of human cognition. In addition to maintaining a persistent external memory, IT can help people to be more
Proposals for Research

reflective and creative – as has been demonstrated in computer support for brainstorming and decision-making (Connolly, 1997; Vogel et al., 1987). However, as our research to date indicates, despite the fact that the Web seems to offer a promising technological base for such a development, computer-supported collaboration is a complex process that requires a sophisticated body of knowledge that we are just beginning to assemble. Moreover, the potential is beyond the reach of any single research group.

We believe that IT support for collaborative knowledge building has not yet been developed to near its potential. KBE research has been carried on for over a decade now, starting with the CSILE system and continuing with KIE, CoVis, etc. (Cuthbert, 1999; Pea, 1993; Scardamalia & Bereiter, 1991; 1996). Recently, commercial systems like KnowledgeForum, WebCT, and LearningSpace are catching on. However, as yet there has been no systematic attempt to support the variety of activities that are involved in knowledge building. There is no general theory of collaborative knowledge building as a social process. Existing research tends to target specific contexts like middle school science with specialized closed systems, rather than developing interchangeable, open source components that can be applied in a full range of contexts. While networks of KBE researchers are coming together in other countries, there is little organized effort to collaborate in the US. The proposed project aims to change this situation. However, collaboration across institutions cannot be started by just wishing for it. This project tries to put some of the necessary conditions in place by developing technical infrastructure (standards, a Perspectives Server, interface components) and initial results that can be used to stimulate discussion and collaboration among KBE researchers locally, nationally, and internationally. Interoperability and collaboration will allow isolated advances to be exchanged, new functionalities to be shared, and test data to be compared.

A Model of Collaborative Knowledge Building

One approach to better understanding how to design computer support for collaborative knowledge building in social settings is to conceptualize the various constituent activities involved in individual and social knowledge building. The diagram below from (Stahl, 2000b) provides a starting point for this, combining aspects of activity theory, situated learning, hermeneutic philosophy, and distributed cognition theory (Chaiklin & Lave, 1993; Cole, 1996; Engeström et al., 1999; Gadamer, 1960/1988; Hutchins, 1996; Lave & Wenger, 1991; Nardi, 1996).
The idea of this diagram is that knowledge building can proceed through many different activities. The sequential structure of the model is only illustrative of an ideal conceptualization. We understand that these activities complexly overlap in practice. The possible relationships among the individual activities – and particularly the interactions between the personal and social – can be complex and varied. The purpose of the diagram is to suggest a number of distinct activities that could be supported by a KBE with multiple functionality. The sequential labeling of these activities corresponds to proposed KBE components listed in Table 1 below, and it is not intended to imply a necessary order to the activities.

A set of seminal books and articles in Computer-Supported Collaborative Learning (CSCL) has formulated a view of learning as a social process of collaborative knowledge building within communities of practice (Brown & Campione, 1994; Brown & Duguid, 1991; Lave, 1991; Lave & Wenger, 1991; Pea, 1993; Scardamalia & Bereiter, 1996; Wenger, 1998). However, these texts do not make the set of cognitive and social activities that underlie such a view explicit in the manner attempted in our KBE theory.

Starting in the lower left corner, Figure 1 shows a cycle of personal understanding. The rest of the diagram depicts how personal beliefs can be articulated in language and become part of social interaction. Note that the results of social knowledge building eventually feed into personal understanding, providing the evolving toolkit of culturally-based individual cognitive capabilities. The depicted knowledge-building activities are discussed briefly below in the context of proposed computer support.

IT Support for Knowledge-Building Activities
Each of the activities of social knowledge building pictured in Figure 1 can be supported computationally. Table 1 lists an illustrative form of support for each. It also lists corresponding prototypes that we have developed. Support for each activity is briefly discussed following the table.

Table 1. Forms of computer support for knowledge building activities.

<table>
<thead>
<tr>
<th>Knowledge-building activities</th>
<th>Forms of computer support</th>
<th>Prototype systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>articulate in words</td>
<td>articulation editor</td>
<td>DynaClass</td>
</tr>
<tr>
<td>public statements</td>
<td>personal Perspective</td>
<td>WebGuide</td>
</tr>
<tr>
<td>other people's public statements</td>
<td>comparison Perspective</td>
<td>WebGuide</td>
</tr>
<tr>
<td>discuss alternatives</td>
<td>discussion forum</td>
<td>DynaClass</td>
</tr>
<tr>
<td>argumentation &amp; rationale</td>
<td>argumentation graph</td>
<td>InfoMap</td>
</tr>
<tr>
<td>clarify meanings</td>
<td>glossary discussion</td>
<td>DynaGloss</td>
</tr>
<tr>
<td>shared understanding</td>
<td>glossary</td>
<td>DynaGloss</td>
</tr>
<tr>
<td>negotiate perspectives</td>
<td>negotiation support</td>
<td>WebGuide</td>
</tr>
<tr>
<td>collaborative knowledge</td>
<td>group Perspective</td>
<td>WebGuide</td>
</tr>
<tr>
<td>formalize and objectify</td>
<td>bibliography discussion</td>
<td>Sources</td>
</tr>
<tr>
<td>cultural artifacts and representations</td>
<td>bibliography</td>
<td>Sources</td>
</tr>
</tbody>
</table>

(a) Computer support should facilitate the process of articulating ideas and preserving them in convenient forms. Most KBEs, including discussion forums like DynaClass, provide an editor for articulating ideas. Some KBEs have tried to introduce procedural facilitation, scaffolding, or prompting to encourage someone to articulate an appropriate expression (Slotta & Linn, 2000). Other approaches would be to provide an outline editor or a brainstorming area.

(b) Public statements by one person confront those of other people. Computer support can represent the different perspectives from which these statements emerge. Perspectives are more general than representations of individuals themselves, because one person can offer statements from multiple perspectives and several people can agree on a common perspective. Perspectives can be related to one another, for instance deriving from a common perspective that they share. Computational representations of perspectives in a KBE like WebGuide (Stahl, 2000a) make explicit the important relationships among personal and group perspectives, as well as providing means for individuals and collaborative teams to articulate their own perspectives.

(c) A KBE with support for Perspectives should provide comparison Perspectives, in which one can view and contrast alternative Perspectives and adopt or adapt ideas from other people's Perspectives. Comparison Perspectives in WebGuide aggregate ideas from various individual and/or
group Perspectives and allow for comparison of them (Boland & Tenkasi, 1995; Stahl, 1999c). Other systems like D3E (Sumner & Buckingham Shum, 1998b) facilitate commentary on documents by other people, such as reviews of journal articles.

(d) The most common element in current KBEs is the discussion forum. This is an asynchronous, interactive communication system like DynaClass that allows people to respond to notes posted by one another. Typically, there is a thread of responses to entered notes, with a tree of divergent opinions. A KBE should go beyond superficial undirected discussion to converge on shared understandings (dePaula, 1998; Guzdial & Turns, 2000; Hewitt & Teplovs, 1999).

(e) Although every note in a discussion forum is a response to another note, the discussion may have a more complex implicit structure. One note might argue for or against another or provide evidence to back up the claim of another note, for instance. Such an argumentation structure can be made explicit and formalized in a representation of the argumentation graph. A component like InfoMap that displays the structure of notes graphically can contribute to participants' meta-level comprehension of their knowledge-building activity, pointing out where additional evidence is needed or where alternatives have not been explored (Buckingham Shum & Hammond, 1994; Donath et al., 1999; Suthers, 1999).

(f) An important requirement for constructing group knowledge is the establishment of shared understanding. This can be fostered by clarifying the meaning of important terms used in various competing claims. A glossary discussion can make explicit how different participants understand the terms they use, as in DynaGloss or DocReview (Hendricksen, 1999).

(g) The glossary discussion should result in a group glossary of the agreed upon definitions of important terms. Such a glossary already represents a form of group knowledge. The glossary is, of course, subject to future debate and emendation; it may make sense to define the glossary as a particular display of information from the glossary discussion (Stahl & Herrmann, 1999).

(h) Perhaps the most delicate phase of knowledge building is negotiation. Computer support of negotiation tends by nature to make explicit the factors entering into the negotiation process. This can be extremely harmful to the subtle processes of persuasion if not done sensitively. On the other hand, negotiation is critical to helping multiple perspectives to converge on shared knowledge. Computer support can provide a useful tool – as long as it is carefully integrated with other social
activities that allow for implicit, culturally established interpersonal interactions (Stahl, 1999b). Group Decision Support Systems (GDSSs) have traditionally been independent systems, not integrated with the broader context of knowledge building (Kraemer & Pinsonneault, 1990; Vogel et al., 1987).

(i) The accumulation of negotiated shared knowledge results in the establishment of a group perspective. Like the alternative individual and team (or subgroup) perspectives, the group perspective may be represented in a KBE. In WebGuide, the content of the group Perspective is inherited into the individual and team Perspectives, because it has been accepted by the group. Individuals can then build on this shared knowledge within their own Perspective and even begin to critique it and start the whole cycle over (Stahl, 1999a).

(j) Shared knowledge can be further formalized. It can be represented in another symbolic system or combined into a more comprehensive system of knowledge (Stahl, 1999c). For instance, in academic research knowledge is incorporated in new classroom lectures, conference presentations, journal articles, and books. The discussion of knowledge that has been compiled into publications can be carried out in a bibliography discussion component of a KBE.

(k) Finally, representations of the new shared knowledge in publications and other cultural artifacts are themselves accepted as part of the established paradigm. Although still subject to occasional criticism, ideas in this form more generally provide part of the accepted base for building future knowledge. In academic circles, an annotated bibliography like Sources might provide a useful KBE component to support this knowledge building activity (Sumner & Buckingham Shum, 1998a).

A KBE goes beyond a single-purpose system – like a simple discussion forum – and supports more than one collaborative knowledge-building activity (Muukkonen et al., 1999). It retains a record of the knowledge that was incrementally collected – unlike common chat, newsgroup, and listserv systems that erase contributions after a short period of time. We hypothesize that it should help people to express their beliefs, to discuss them with others, to differentiate their own perspectives and adopt those of other people, clarify disagreements or misunderstandings, critique and explicate claims, negotiate shared understandings or agreements, and formulate knowledge in a lasting representation. Because KBEs are computational, they can provide facilities like searching, browsing, filtering, tailoring, and linking in order to group related ideas together automatically. KBEs can interface with other agents and software utilities – for instance sending
emails to notify collaborators when important knowledge-building events have taken place (McLean, 1999). They can also dynamically format sets of notes in convenient displays for different purposes.

2.3. An Infrastructure for KBEs

Computational Perspectives

Computational Perspectives provide a new, dynamic, personalized form of on-line information management (Stahl, 1995). A Perspective defines an electronic workspace in which a person or group can develop ideas and manage information that belongs together – for instance because it represents the beliefs and viewpoint of a particular person, group, domain, or intellectual position. Perspectives structure a shared information space so that special coherent views can be built up and displayed. Although the mechanism of computational Perspectives is very general and flexible, the simplest way to use it in a small group is to define a personal Perspective for each member, one team Perspective for agreed upon ideas, and a comparison Perspective that collects the ideas from all the personal Perspectives.

The design philosophy behind computational Perspectives as implemented by the PI in WebGuide is that users have complete control over the content in their personal Perspectives. Thus, if my personal Perspective inherits conflicting ideas from different team Perspectives that I belong to, I can delete, edit, and rearrange those ideas at will. Other users can view the contents of my personal Perspective (except for content that I have designated as private) and they can copy items, link to them, initiate public discussions of them, and propose them for incorporation in team Perspectives – but none of this affects how the content of my Perspective is displayed to me. This allows me to build my own Perspective on the topics that are under consideration by the group. I can see what knowledge others are building, incorporate that knowledge into my Perspective, or join in with others to share, discuss, and negotiate. The same design philosophy applies of course to team Perspectives: team members jointly (through negotiation processes) have complete control over the content of their team Perspective.

Inheritance is a central defining mechanism of computational Perspectives as used in this proposal. The ability to define arbitrarily complex networks of Perspectives with multiple layers of sub-groups between the group Perspective and the individual personal Perspectives, and to have the automatic inheritance of content through the network distinguishes this approach from all other systems of “views” and “perspectives.” Inheritance
in this sense is not class inheritance, but “content inheritance.” A given Perspective can inherit content from multiple other Perspectives. This content is aggregated (logical union) in the given Perspective, where it can be over-ridden with edits, deletions, rearrangements, and additions. The inheritance mechanism is derived from efficient approaches explored in hypermedia, including “delta memory” and “transclusion” (Boborow & Goldstein, 1980; McCall et al., 1990; Mittal et al., 1986; Nelson, 1981; Nelson, 1995). For a discussion of related work, see (Stahl & Herrmann, 1999).

Because new Perspectives can be defined (either in advance or during system use) to inherit from any (non-cyclical) other Perspectives, it is generally useful to define “comparison Perspectives” that aggregate the ideas from team members, including those ideas that have not been agreed upon and migrated to the team Perspective. This is handy for keeping an eye on what one’s fellow team members are thinking. Typically, we have set up the inheritance network of Perspectives to have a diamond-shaped profile, diverging out from the total group Perspective via teams to all the personal Perspectives, and then converging back via team comparisons to the group comparison Perspective. This models a collaborative knowledge-building process that combines divergent brainstorming and convergent negotiation.

Functionally considered, a KBE with Perspectives like WebGuide consists of two primary subsystems: a Perspectival data selection computation and a set of interface displays of the selected data. When a display is requested, the system must search the database to determine which content notes should be displayed to the particular user in the requested Perspective. For instance, if I request to view your Perspective, the system must select notes that are defined within your Perspective or within any Perspective from which yours inherits (recursively), except for notes that are private or that have been over-ridden. Various special displays can also be computed using this inheritance computation by treating discussions, negotiations, historical archives, etc. as pseudo-Perspectives that have special inheritance and exclusion rules. Once the Perspectival data computation has been returned, the content can be displayed in specialized interfaces that provide different kinds of functionality useful for further knowledge building.

**An Open Source Perspectives Server**

A specific task of the proposed project is to separate out the Perspective computation from WebGuide and structure it as a self-contained module with a well-defined application programming interface (API). This will form a Perspectives Server, a Java application that runs on the Web server along
with the database system. It will be separate from the WebGuide client that will still run in a Web browser on the client’s computer. This separation of functions into a server and a client will have many advantages. It will speed the functioning of WebGuide because the intensive computation of Perspective content will be done on a central server that is faster than typical student computers. Also, calls to the database system will take place locally rather than across the Internet. In terms of system development, it will mean that developers can build systems that incorporate Perspectives without having to worry about the Perspective algorithms or the database calls. They will use an API that lets them request data that should be shown to a given user in a given Perspective. They can then just focus on how best to display this data in the interface.

The Perspectives Server will be a self-contained Java application. It will be released as open source with clear documentation on how to use it to get Perspective data for display. The data will be delivered as an XML text stream that can be used by any Web technology, such as HTML, Perl, or Java. The data will be human-readable, making it easy for programmers to see what data is being passed. Although it is anticipated that the Perspectives Server will generally be used as a black box, its open source availability will allow programmers to modify it if necessary, such as to incorporate improvements to the XML DTD or in response to changes in Web technology. However, the Perspectives Server will be designed to make expansions of the database schema easy to incorporate without changes to the source code. This will allow new data structures corresponding to new multimedia data types.

The Perspectives Server will be a form of middle-ware, operating between the database and the client software. It will instantiate a three-tier, model-controller-view architecture that defines independent layers for the data schema or model, the data computation or control, and the interface display or view. The database management system can be any standard relational SQL system like mySQL or Oracle. The middle layer can be the Perspectives Server or a stripped down version that does not compute Perspectives. And the interface can be any kind of applet, Web page, or Web application that conforms to the API standard.

A Component Architecture for KBEs

The release of the open source Perspectives Server will not only facilitate the rapid prototyping of Perspectives-based systems for this project to use in its study sites, it will also allow other researchers to incorporate
Proposals for Research

computational Perspectives in their KBEs. We have already had requests for this from researchers in California and in Germany.

The PI of this proposal is involved in several efforts to promote collaboration among KBE researchers. Among these, he is the PI on a seed grant from the NSF-funded Center for Innovative Learning Technologies (CILT) to foster data interoperability among KBE systems. Work on this is currently producing a draft XML DTD (eXtensible Markup Language – Document Type Definition) to provide a common data format that KBE data can be imported to and exported from. Data in this format can be displayed using CSS and XSL. Tools we are now developing will allow such data from any KBE to be analyzed with standard measures, allowing for instance the volume and characteristics of discussion threads on different systems to be quantified and compared.

This DTD will also provide the format for data transfer between the Perspectives Server and interface clients. Clearly, interface clients will have to be designed to accept and make use of data received in this XML format. An important aspect of the proposed work will be to define a set of standards: the XML DTD for data interchange, the API for the Perspectives Server, and the ability of interface clients to call the Server and make use of the data.

The PI will be communicating with other KBE researchers nationally and internationally to solicit their concerns about these standards and to work toward a consensus and adoption of a set of such standards. These standards do not have to be formally approved by international standards bodies; informal agreement within a set of collaborating research groups is all that is needed for substantial practical benefits. Use of these standards will allow for rapid prototyping and customization of systems for various study sites both within the proposed project and by other researchers. It will also move us significantly toward a future in which KBE components from different research groups can be intermixed so that new systems can take advantage of functionality developed at different sites. Certainly, every effort will be made to incorporate related international standards, such as those for XML, XLINK, and metadata.

2.4. Study Sites for Evaluating Knowledge Building in Workgroups

We will use local sites under our own control as alpha sites for testing our software, not only to eliminate bugs, but to try out different functionality and to refine the interface. More formal evaluation of the social impact of KBEs in workgroups will be conducted in corporate sites, primarily in the
Boulder area. We have targeted two local situations (academic research and university learning) and two corporate sites (corporate training and industrial design).

**Academic Research: the L3D Center**

The proposed work will take place within the Center for LifeLong Learning and Design (L3D), a research group within the Department of Computer Science at the University of Colorado. We will try out our prototypes in a variety of applications within L3D. Such self-application will give us first-hand experience with the requirements for the use of KBE software and with the practical problems of deployment and adoption. Increasingly, research at L3D involves participants from different disciplines and even virtual subgroups, like colleagues at other universities and other countries. When, e.g., a research project involves participants from Boulder, Colorado, and from Dortmund, Germany, a Web-based collaboration medium is essential, and means for defining and negotiating personal, subgroup, and whole-group Perspectives seem particularly appropriate. Because users in this group are themselves software developers and researchers, they are particularly accepting of glitches and are reflective about design issues. This will provide a convenient and forgiving initial test site.

**University Learning: College Seminars**

Several members of L3D, including the PIs, offer undergraduate classes and graduate seminars. These courses typically emphasize student on-line discussion and group projects. They often have a content focus on the Internet. The PIs have used WebGuide and other Web-based tools in past courses and will continue to use them in the future. We are exploring courses that not only span multiple disciplines, but also span the oceans. Again, here, Perspectives for subgroups (interest-, content-, discipline-, or location-based) make sense. Collaborative classrooms will provide a secondary test site in which KBE functionality and prototypes can be tried out and knowledge-building activities can be monitored.

**Corporate Training: Athenaeum International DesignShops**

We foresee our primary study site for this project being a corporate training setting, in which knowledge building takes place under settings that may be advantageous for study. Athenaeum International (AI) is part of a distributed network of corporate training facilitators associated with MG Taylor. AI is located in Boulder and has established a good working relationship with the
Pi and L3D. AI specializes in the design and manufacture of custom movable furniture for rapid deployment at corporate training events, as well as the facilitation of such events. They are interested in incorporating knowledge-building software tools in their furniture to support the training process.

A typical training event – or DesignShop™ – involves bringing together decision-makers from throughout a company to “reengineer” their corporation or re-think their high-level mission. This might involve a series of three-day workshops, or even an on-going sequence of quarterly gatherings. AI staff have noted a number of problems that they think could be addressed by innovative computer support: there is tremendous time pressure and everyone cannot express all their ideas and arguments; too much time is spent introducing materials; it is hard to retain important points and decisions; follow-through is tricky; documentation is labor-intensive. KBE support could include pre-workshop preparatory discussions, capturing of ideas that arise in the face-to-face meetings, organized documentation of debates that took place, and follow-up discussion, analysis, negotiation, decision-making, or follow-through.

AI would provide a challenging study site for KBEs. DesignShops are high-stakes events involving people who need to make efficient use of their time. The groups here would be larger than the test groups at the university, and would involve more intense face-to-face interaction. A workshop series is of limited duration, so success could be assessed quickly and changes made prior to a subsequent trial. Project staff would have the aid of AI’s experienced group process facilitators to guide the design of the software support and of the deployment strategy, as well as to analyze the impact the software had on the social systems.

**Industrial Design: Seagate Technical Design Centers**

Another targeted corporate study site provides a rather different opportunity for investigating the use of KBEs. We have begun to investigate a particular work group within Seagate, a major hard disk manufacturer located near our university. Seagate is an established high-tech corporation. Through a series of acquisitions and mergers, it now consists of large design centers in Longmont (outside of Boulder, Colorado), Minneapolis (MN), Kansas City (IW), Redwood City (CA) and Singapore. Each of these design centers houses a few hundred employees working rather independently of the other centers.
We will focus on the effort of a Seagate Vice-President who is in charge of coordinating research on the problems of disk drive head tracking. In order to meet market demands that are projected as exceeding Moore’s law — requiring a doubling of storage density every year — the designers who work on head tracking must solve complex issues in physics and mechanical engineering. Unfortunately, engineers at different centers work almost independently of each other, duplicating research and designing products that overlap in functionality and specifications. As a result of their different histories, the centers have very different cultures of work, interaction, and outlook. For instance, one center prides itself in minimizing costs while another spares little cost to make what they consider a quality product. Designers from the different centers are accustomed to different engineering paradigms and find it hard to talk to each other.

The research objective here would likely be to use KBEs to structure communication and collaborative design among the distributed design groups. There is already a shared Lotus Notes database in which everyone can view the specifications and deadlines for each group’s product line. However, there is currently no medium of communication among the groups (other than generic email) and no persistent textual discussion of the posted data. This makes it hard to share interpretations, work on establishing common understandings, or collaborate on building knowledge from the data. There is no support for Perspectives that would represent the conflicting cultures of the distributed groups and allow for negotiation of these differences.

Other Potential Study Sites

We will try to work with both AI and Seagate initially to explore their suitability to this project. We may end up focusing on one or the other site for practical or theoretical reasons. It is always hard to predict how field research will develop over years, particularly in today’s volatile marketplace, where key contacts change jobs and companies alter their strategies. For this reason, we have also established access to several other industrial sites similar to AI and Seagate: StorageTech is another major manufacturer of digital storage media; our contacts there are especially concerned with making their company a “learning organization” in which knowledge building is recognized to be an essential aspect of work. IBM has a major support center near Boulder, with a help-desk organization of 700 employees who must continually collaborate to build knowledge of the products they support. L3D has maintained a long-term relationship with two Japanese software companies, SRA and PFU, who are both interested in
incorporating our ideas about lifelong learning and KBEs into their operations. These companies all have interesting settings where we would be welcome to deploy and observe our system prototypes if we have time during the proposed project.

Evaluation of Social Impact of KBEs at Study Sites

Evaluation will be conducted using converging methods to understand the complex, systemic issues around new technology deployment and use. Evaluation goals are two-fold:

- **Constructive**: We want to understand the environment with the objective of tailoring the design of the KBE technology to the study site, as well as constructively guide the deployment toward successful adoption.

- **Objective**: We want to objectively observe evolving use of the KBE technology; assess the nature of the IT impacts on coordination and collaboration; and evaluate the validity of the "collaborative knowledge building" concept and applicability of the KBE theory; and refine the KBE theory based on results from observed practice.

Data collection will take the form of face-to-face in-depth interviews; phone and email-based "interviews" when subjects are at a distance (particularly in the case of Seagate); real-time non-participant observation in the workplace as well as via the KBE technology; surveys (particularly in the case of the AI training sessions, where there will be many more subjects); KBE database data collection; and document collection where appropriate.

In addition to coding and content analysis of field and interview notes, analytical approaches will also include structural analysis of discussion thread lengths and participation levels based on the KBE database data. We will perform discourse and content analysis of argumentative exchanges, and KBE-captured collaboration episodes.

Issues we will investigate include:

- What is the nature of collaborative knowledge building, and what activities comprise it?

- How does asynchronous support for articulation affect participation with respect to a variety of factors including time, location, and social status?

- How does computational support for discussion, argumentation, and clarification affect consensus-building as well as conflict?
• Do subjects understand and put into practice the concept of Computational Perspectives? How do subjects interact with and manage multiple Perspectives (personal, subgroup, group, comparison)?

• Can "synergistic moments" be captured? What portion of these moments occur on-line in KBE environments?

• What do subjects do with persistent dialog? Does persistence foster reflection on "synergistic moments," and under what circumstances (what shades of consensus and conflict)?

• What role do facilitators (in the case of Athenaeum International) and management play in the use of KBE technologies?

• What are the particular hurdles that must be overcome for KBEs to be useful in these particular organizational environments?

Since the evaluation is a multiple-person effort, and because there will likely be different researchers participating at different times during the pre- and post-deployment stage, we will carefully organize observation efforts, and make an effort to systemize field note format to the best possible extent. We will also conduct regular group data analysis meetings to coordinate the results of our efforts.

3. OBJECTIVES OF PROPOSED WORK

3.1. Objectives for Theory Development

• Investigate the phenomenon of the "synergistic moment" in interactions captured in the database.

• Investigate the utility and actual usage of computational Perspectives.

• Refine the model of collaborative knowledge-building activities based on project findings.

• Publish results of this project in a monograph.

3.2. Objectives for Technology Prototyping

• Define a standard for data interoperability among KBEs.

• Release open source import/export/display/analysis tools for KBE data interoperability.
Proposals for Research

• Release an open source Perspectives Server.
• Release example KBE interface components using different technologies.

3.3. Objectives for Deployment
• Deploy a KBE prototype in settings of academic research (e.g., L3D research group).
• Deploy a KBE prototype in settings of university learning (e.g., seminars at the university).
• Deploy a KBE prototype in settings of corporate training (e.g., AI DesignShops).
• Deploy a KBE prototype in settings of industrial design (e.g., Seagate design groups).

3.4. Objectives for Evaluation
• Conduct initial evaluation of the way work is presently conducted among selected groups to: collect baseline data; carefully select groups to deploy to; and formulate deployment strategies.
• Assess existing challenges for communication and collaboration within and across groups through interviews and possibly email-based interview-type surveys for those participants at a distance.
• Observe deployment of KBEs and revise deployment strategies as necessary.
• Perform on-going qualitative (observation, interviews, and document collection) and quantitative (KBE database activity) data collection.
• Analyze collected data to evaluate the impact of the KBEs in the workgroups.

3.5. Objectives for Dissemination
• Build a local, interdisciplinary community of students and faculty to conduct KBE research.
• Collaborate with at least 3 US research groups and with at least 3 international KBE research networks.
• Maintain an active website with the results of this project, including the open source products.

• Present the work of this project in the Group ’01 and CSCL ’01 international conferences (to be held at the University of Colorado), and report the findings of this project at 3 or more other international conferences.

4. PLAN OF PROPOSED WORK

4.1. Year I

Refine the model of collaborative knowledge-building activities through presentations to researchers and analysis of the "synergistic moment" in face-to-face collaboration.

Define a standard for data interoperability among KBEs using XML and XLINK for notes and relations among notes, including threaded discussions; explore the adequacy of this standard using local prototypes; circulate the proposed standard among other KBE researchers; integrate the proposed standard with emerging data standards.

Define a standard for KBE interface components including JavaBeans, using XML for data interchange; explore the adequacy of this standard using local prototypes; circulate the proposed standard among other KBE researchers; integrate the proposed standard with other emerging data standards.

Define a standard for KBE database servers to provide XML data structures to interface components meeting the above standards.

Release an open source Perspectives Server that efficiently computes data visible in a requested Perspective in the standard XML format. Provide open source and documentation so that other researchers can use this Server for Perspectives-based KBEs and can propose improvements to the Server.

Deploy a KBE prototype in a setting of academic research such as a local research group. The KBE will be used to discuss and design standards and software. Learn from this deployment experience.

Begin initial observation of two primary test sites (AI and Seagate) to prepare for eventual KBE deployment.
Collaborate with American research groups interested in KBEs (e.g., Berkeley, SIU, Stanford, Georgia Tech, Hawaii, SRI) and international KBE research networks (e.g., existing CSCL research networks in Germany, England, Norway, Finland, Canada) to coordinate software component design around agreed upon standards.

Establish a website to publicize the results of this project, including the open source products, and to provide an information center for collaborations. Incorporate a KBE in the website to foster interactive knowledge building related to the proposed work.

4.2. Year II

Release open source import/export/display/analysis tools for KBE data interoperability, using XML to define a DTD and using XSL and CSS to display the data. Document the standard and the tools on a website that makes them available to other researchers.

Release example KBE interface components in different technologies, including Java, Perl, and HTML. Make these available on a website with documentation and tutorials to help other researchers develop compatible components.

Deploy a KBE prototype in a setting of university learning such as a seminar offered by the PI, and observe use. The KBE will be used for students to develop, exchange, and negotiate reflections on shared readings and on collaborative writing projects.

Continue investigating the AI study site and assess the requirements for deployment of a KBE there.

Deploy a KBE prototype in a setting of corporate training such as a series of AI corporate training sessions. Commence post-deployment observation.

Present approach and intermediate findings of this project at the Group ’01 and CSCL ’01 international conferences. During Year II, these conferences will both be hosted at CU, with the PI serving as chair of the local arrangements committee at Group and chair of the program committee at CSCL.

4.3. Year III

Deploy a KBE prototype in a setting of industrial working such as a technical design group at Seagate, and observe and analyze use there. The AI deployment will also continue to be evaluated in Year III.
Continue on-going analyses of KBE use and examine findings with respect to the project's central hypotheses about KBEs. Publish a monograph on the project findings.

Report on this research at 3 or more international conferences.

**5. EXPECTED IMPACT OF PROPOSED WORK**

The proposed work is designed to provide some of the basic conditions for the sorts of intensive collaboration that we believe are necessary to achieve the potential of KBE technology. In addition to providing technical conditions (e.g., Server, architecture, standards) and social conditions (local, national, international networks), the work will prepare specific examples of functionality (e.g., computational Perspectives) and concrete analyses of application. These results will be disseminated through communication channels including specific related conferences.

Within the PIs’ home institutions the proposed work will be an important catalyst for building an interdisciplinary collaborative effort of undergraduates, graduate students, research staff, and faculty working on KBE theory, technology, and evaluation. The PI and his colleagues have already begun to involve education, communication, and computer science students in this work through seminars, independent study, and class projects. The proposed work defines a coherent research agenda over several years that will crystallize a local research community.

**REFERENCES**


transclusion, Communications of the ACM, 38 (8), pp. 3-33.
Norman, D. A. (1993) Things That Make Us Smart, Addison-Wesley
Publishing Company, Reading, MA.
In G. Bjerknes, P. Ehn, & M. Kyng (Eds.), Computers and Democracy:
http://seed.cs.colorado.edu/dynasites.documentation.fcgi.
Palen, L. (1999) Social, individual & technological issues for groupware
calendar systems, In: Proceedings of ACM Conference on Human
Basic Books, New York, NY.
Pea, R. (1993) The collaborative visualization project, Communications of
the ACM, 36 (5), pp. 60-63.
cooperation as opportunities for learning. In W. P. Robinson (Ed.)
203-234.
Peters, J. D. (1998) Speaking Into the Air, University of Chicago Press,
Chicago, IL.
Scardamalia, M. & Bereiter, C. (1991) Higher levels of agency in
knowledge building: A challenge for the design of new knowledge
building communities. In T. Koschmann (Ed.) CSCL: Theory and
Practice of an Emerging Paradigm, Lawrence Erlbaum Associates,
Slotta, J. D. & Linn, M. C. (2000) The Knowledge Integration Environment:
Helping students use the Internet effectively. In M. J. Jacobson & R.
Kozma (Eds.), Learning the Sciences of the 21st Century, Lawrence
Erlbaum Associates, Hillsdale, NJ.
Stahl, G. (1975) Marxian Hermeneutics and Heideggerian Social Theory:
Interpreting and Transforming Our World, Ph.D. Dissertation,
Department of Philosophy, Northwestern University, Evanston, IL.
Available at:
m.


Available at:
http://www.cs.colorado.edu/~gerry/publications/journals/ai&society/.

Stahl, G. (2000b) A model of collaborative knowledge building, In:
Proceedings of International Conference of the Learning Sciences (ICLS 2000), Ann Arbor, MI. Available at:

Stahl, G. & Herrmann, T. (1999) Intertwining perspectives and negotiation,
In: Proceedings of International Conference on Supporting Group
Work (Group ’99), Phoenix, AZ. Available at:

Stahl, G., Sumner, T., & Owen, R. (1995) Share globally, adapt locally:
Software to create and distribute student-centered curriculum,
Computers and Education. Special Issue on Education and the Internet,
24 (3), pp. 237-246. Available at:
http://www.cs.colorado.edu/~gerry/publications/journals/c&e/.

Sumner, T. & Buckingham Shum, S. (1998a) From documents to discourse:
Shifting conceptions of scholarly publishing, In: Proceedings of
Computer-Human Interaction (CHI 98), Los Angeles, CA, pp. 95-102.

Sumner, T. & Buckingham Shum, S. (1998b) A toolkit for publishing Web
discussion documents: Design principles and case studies, In:
Proceedings of Asia Pacific Computer Human Interaction (APCHI 98),
Shonan Village Center, Japan, pp. 218-223.

Suthers, D. (1999) Representational support for collaborative inquiry, In:
Proceedings of 32nd Hawai’i International Conference on the System
Sciences (HICSS-32), Maui, HA, pp. CD-ROM. Available at:

decision support systems: Determinants of success, In: Proceedings of

Cambridge, MA.

Kaufman, Los Altos, CA.

Wenger, E. (1998) Communities of Practice, Cambridge University Press,
Boston, MA.

A New Foundation of Design, Addison-Wesley, Reading, MA.
Models for Organizing Collaboration: Ways of Supporting Distributed Learning

Proposal to Lotus Research

Primary Research Objective

The goal of this research is to identify a set of models of collaboration that can serve to guide both (a) the design of collaboration software by Lotus and (b) the application of this software to specific learning situations by user organizations. The identified models will be compiled and presented in a format that has been demonstrated to be usable and useful in supporting distributed learning.

The following sequence of questions will be investigated:

• How do people learn what they need to know as part of their collaborative work?

• What are the major phases of collaborative knowledge building according to current theories?

• What are effective instructional methods for promoting distributed learning according to current best practices?

• What forms of computer support can support these knowledge-building phases and instructional methods?

• How can user organizations be guided in organizing the functionality of Lotus software and other resources to promote collaboration?

These questions will be pursued from the perspective of informing on-going planning, design and research on collaboration and distributed learning software at Lotus. It will build upon the existing expertise and research of
the project partners and will be prioritized to meet stringent project time constraints.

**Partners**

*Researcher:*

Gerry Stahl, University of Colorado
Jose Rafael Lopez Islas, Monterrey Tech Inst
Kamran M. Khan, Marist College

*Lotus Representatives:*

Randy Cox, Director of Engineering
Nada Abu-Ghaida, Product Designer

The researchers each bring to this project a background in relevant academic research and working relationships with practitioners of distributed learning. The Lotus representatives come from strategic positions within the Lotus software development process and also bring working relationships with practitioners who will help to evaluate the results of this project. (See attached resumes.)

**Gerry Stahl** is a Research Professor at the Center for LifeLong Learning and Design of the University of Colorado, with a joint appointment in Cognitive Science and Computer Science. His specialty is the theory and design of collaborative knowledge-building environments.

**Jose Rafael Lopez Islas** is the Director of Research and Educational Technology at the Monterrey Institute of Technology's Virtual University. The Virtual University enrolls students throughout Latin America and the Monterrey Tech System is the largest user of Lotus' LearningSpace. Professor Lopez' research focuses on the social construction of knowledge.

**Kamran Khan** is the Vice President for Information Technology and Chief Information Officer of Marist College. His specialty is Distributed Learning, Knowledge Management and E-Commerce in education and corporate partnerships.

**Randy Cox** is Director of Engineering for Lotus at Redwood City, where he leads work on the next version of LearningSpace. He proposed this project
on models of collaboration in order to provide a theoretical framework for planning future Lotus software.

Nada Abu-Ghaida is a Product Designer at the Cambridge offices of Lotus, where she designs interfaces for future Lotus learning technologies. She agrees with the importance of this project and emphasizes the need to help user groups to organize software functionality and other resources to support distributed learning

Proposed Scope of Work

This project addresses the problem of how to organize software functionality and content to support distributed learning. Instructors of distributed training and classes have few guidelines for how best to support distributed collaborative learning using new computer-based technologies. For instance, how does one organize course materials and activities including readings, simulations, collaborative projects, group discussions, negotiation processes and portfolio artifacts into an effective educational experience that achieves targeted instructional objectives? Even designers of this technology have no place to turn for an overview of current theoretical frameworks and best practices that can inform their designing of future functionality. If Lotus is to maintain its leadership position, it must go beyond generic discussion software and course administration with innovative functionality to support multiple phases of collaboration and of the social (i.e., group and organizational) construction of knowledge. In addition, it must provide guidance to its user community on how to organize computational and digital resources for effective distributed learning.

Stage I

The objective of this project is to develop and test a methodology or a coherent set of principles and practices for organizing software functionality to support distributed learning. This will be based on a review of models of knowledge-building activities. Stage I of the project will be the systematic review of models of collaboration from the perspective of informing computer support of distributed learning. These models will draw on several of the most influential theories (e.g., situated learning, activity theory, constructivism) and instructional design approaches (e.g., case-based and problem-based learning). The project will either develop a model that combines ideas from these different theories or will compile a set of several
models of collaborative learning that are applicable to computer-supported distributed learning.

The project participants are all experts in the topic of the project at both theoretical and practical levels. Thus, they are already familiar with many of the important theoretical approaches and are experienced in the issues of distributed classroom and training settings. The project will bring together this existing knowledge, fill in important missing areas of knowledge, and organize the knowledge in a coherent and useful way.

Perhaps the most important work product is the research survey. We will attempt to accomplish this prior to summer (March - May).

In early summer, we will have a project meeting/workshop where we evaluate our findings. If we feel that we have substantially completed the research phase, we will proceed to the development and testing of a guidebook based on our research. If we feel it is important to continue the research work, we will revise our project schedule in order to ensure that the research phase is adequately accomplished. The goal is not to exhaustively catalog every possible theory, but to make sure that we have gleaned the most important implications of the major theories and methodologies. The research phase will culminate in a digital catalog of the models we have found and developed. This will include theoretical models of collaboration and individual elements in the collaborative process, relevant theories of cognition and of instructional design, existing best practices, and current commercial support tools.

**Stage II**

Following the research on models for organizing collaboration and our first project workshop, we will develop a guidebook for designers of distributed learning based on these models. This will be a practical guide with principles and step-by-step procedures for organizing software tools and instructional content into effective and coherent presentations. We will try to develop this methodology and draft the guidebook over the summer (June - August). Marist College is a Lotus shop and has staff and students trained in LearningSpace and other Lotus products; Khan and colleagues will take a lead role in implementing the guidebook.

**Stage III**

From July - November, we will assess the usability and effectiveness of the methodology and guidebook in business, research and educational settings. This will not involve the development of any new technology. Ideas for new
software functionality will be assessed through mock-ups, scenario walk-throughs and in-depth surveys. The guidebook will be reviewed by potential users and will be revised in response to their feedback. The academic researchers will use typical situations in their own institutions and activities to test the guidebook in academic distributed collaborative learning settings. They will also use their industry contacts to involve industry representatives in reviewing the guidebook from the perspective of industrial distributed collaborative learning settings. For instance, Stahl will work with corporate trainers at Athenaeum International, with managers of distributed research groups at Seagate and with employee development trainers at StorageTech, to assess the guidebook's applicability to a wide range of industrial distributed learning tasks. Lopez will test the guidebook in graduate courses taught through the Virtual University, with students located in different countries of Latin America and all over Mexico. The Lotus representatives in the project will work with people from their alpha and beta test sites and from their user community to conduct similar assessments of the guidebook.

**Deliverables and Timeline**

Contract and project start and end dates: March 1, 2000 - December 31, 2000.

**Stage I**
March - May: Catalog of models of collaboration and distributed learning instruction. The catalog will be deployed in Lotus Notes/Domino. First workshop of project participants.

**Stage II**
June - August: First draft of guidebook on how to support distributed learning. The guidebook will be deployed in Lotus Notes/Domino.

**Stage III**
July - November: Review of guidebook by user organizations. Publication of revised guidebook in Lotus Notes/Domino. Second workshop of project participants and possible third workshop with user organizations and/or Lotus representatives.

December: Final report and wrap-up.
POW! (Perspectives On the Web)

Proposal to Colorado Advanced Software Institute (CASI)

The asynchronous management and negotiation of knowledge in shared collaboration spaces should be supported by software that keeps track of personal, group and comparison perspectives. The POW! Project will produce a Java server to facilitate quick development of Web-based client software incorporating perspectives to support collaboration in educational and corporate settings. The POW! Project will release the Java server on the Web with an appropriate license. It will have a clear API, XML data exchange and sample code for educational clients in Java, HTML and Perl. The Project will also develop requirements for corporate applications in management and training.

Problem, Background and Opportunity

We are speeding toward a society in which people are networked together to share information and to learn and work collaboratively. The hardware infrastructure is developing rapidly, with Colorado in the communications forefront. Technologies for simple exchange of information _ like email and audio/video/textual conferencing _ are being widely adopted. However, software support for the collaborative construction of deeper knowledge remains an open research issue.

Teamwork, flexibility and collaboration are becoming the mode of operation for modern companies, whose employees may be geographically dispersed. Companies must make complex decisions that synthesize the expertise of many employees; they must become _learning organizations_ that share effective, evolving _organizational memories_. Imagine a corporate reengineering workshop in which people from throughout an organization gather (physically and/or virtually) to set a new strategic direction: how could software support this process by facilitating the construction, sharing
and synthesis of different perspectives on the problem? The design of such software goes far beyond what is available today and involves consideration of both technical and social issues.

If we can develop sophisticated conferencing software for Colorado companies to meet their own organizational learning and decision-making needs and to provide training for other corporations, then we will complement Colorado’s strengths in the hardware and communications sectors and move into a leadership position in educational and groupware research.

**Objectives**

The PI has developed a research prototype named WEBGUIDE designed to support deep knowledge construction by collaborative groups over the Web. Testing in classroom situations has suggested several tasks needed to make this software practical for transfer to industry:

- to increase WEBGUIDE’s speed and flexible further development or application.
- to explore its use in corporate settings of management and training.
- to make it available to other researchers to collaborate on further development and assessment.
- to allow for free development of alternative interfaces for different applications.

To accomplish these tasks, the POW! Project has the following objectives:

- to separate WEBGUIDE into: (i) a perspectives server that carries out the intensive computation of perspectives and sends requested data in XML format to (ii) light-weight clients.
- to study application of WEBGUIDE in corporate decision-making and corporate training situations.
- to release the POW! perspectives server as Open Source Software (OSS) under a license acceptable to the University and the Collaborating Company in order to encourage use of the server by other researchers.
- to develop and document sample light-weight clients for educational applications using Java, HTML and Perl technologies to demonstrate
how Colorado corporations can quickly develop proprietary clients for corporate applications using the POW! perspectives server.

Accordingly, the POW! Project has the following milestones and deliverables:

- by 1st quarter of grant: to create a POW! perspectives server in Java with a clear API and with XML data exchange to Java, HTML and Perl clients. Optimize, modularize and document the server code and API.

- by 2nd quarter of grant: to negotiate an OSS license with the University and the Collaborating Company and to release the POW! server under this license on a website with appropriate documentation to support collaborative development of the software.

- by 3rd quarter of grant: to develop and document on the website sample clients for the POW! server illustrating client development using alternative technologies.

- by 4th quarter of grant: to study applications of the software within the collaborating company and at its corporate training sessions, resulting in requirements for client applications in these settings.

**Potential for Broad-based Technology Transfer**

Over the past decade the PI has developed a perspectives mechanism to support collaborative knowledge building. He has applied this technology to a number of applications and this work has been widely accepted in the peer-review research community. The WEBGUIDE prototype to be used in the POW! Project implements the perspectives mechanism on the Web and has been assessed in educational contexts.

The Collaborating Company, Athenaeum International, is part of a national network of corporate trainers, MG Taylor, that has extensive experience conducting workshops and design sessions for Fortune 500 companies and other major clients. They use a successful training methodology and are interested in incorporating computer support into their approach.

The POW! Project will synthesize the expertise of both participants through joint planning of software for corporate applications. This will include attendance of the PI and graduate student at various corporate meetings and events; involvement of both sides in collaborative requirements planning for software; and joint assessment of the software in corporate settings.
The POW! Project will significantly further the development of the perspectives-based software for both educational and corporate applications. The server and the sample educational clients will be available under license for companies throughout Colorado and for university researchers to use. All corporate client software developed within the POW! Project will be available under the standard CASI conditions and licensing options.

**Approach**

The POW! Project builds upon successful research by the PI in the past to develop effective support for collaboration in corporate settings. The key innovative technology, a flexible perspectives mechanism was originally developed in 1991-1993 under CASI support and was subsequently used in NASA applications by Johnson Engineering. More recently, it has been implemented on the Web and tested in educational settings. In the POW! Project, it will be re-implemented in an architecture that will facilitate its deployment in corporate settings.

**Past Work by PI:**

As a graduate student working with Professor Raymond McCall, the PI developed a _perspectives mechanism_ within the PHIDIAS hypermedia system (Stahl, 1991; Stahl, 1992; Stahl et al., 1992). This work was supported by CASI grants in 1991, 1992, 1993 _earning CASI_s exemplary Research Award in 1993. The perspectives mechanism was a central part of the PI_s Ph.D. dissertation (Stahl, 1993a; Stahl, 1993b; Stahl et al., 1993a; Stahl et al., 1993b). Since then, the PI adapted the perspectives mechanism to several different application areas, including Hermes 2.0_ and a system for ISO 9000 documentation which the PI developed within his own company (Stahl, 1995; Stahl, 1996; Stahl et al., 1995a; Stahl et al., 1995b). Most recently, the PI developed WEBGUIDE, a Web-based hypermedia educational environment to support collaborative classroom learning, and tested it in both middle school and graduate level classrooms (Stahl, 1999a; Stahl, 1999b; Stahl, 1999c; Stahl & Herrmann, 1999; Stahl et al., 1999). The PI developed a theoretical framework for perspectives and collaboration in his doctoral dissertation and in recent publications (Stahl, 1993a; Stahl, 1999d; Stahl, 2000).

The PI is currently a Research Professor in Computer Science and Cognitive Science and a faculty member of the Center for LifeLong Learning and
Design at CU. He has published widely on knowledge-building software environments, organized a well-attended international workshop on this topic and taught a series of advanced seminars on it. He has developed software since the mid-1960s and has worked with the Graduate Student on WEBGUIDE for more than two years.

**Supporting Collaboration**

Collaboration is an important but difficult and poorly understood activity. The potential is that the ideas, expertise and critical abilities of a number of people can be synthesized to produce knowledge that no one participant could have produced and to share this knowledge among all participants. Software can support this process by providing an external memory or workspace in which each participant can develop personal ideas, can view the ideas of others, can incorporate others’ perspectives into their personal perspective and can negotiate agreements and clarify points of difference within the group as a whole. A computer-based environment can maintain persistent views of ideas that have been expressed, so that one can review the history of discussions and compare related ideas. A Web-based system can facilitate collaboration among people who are not present at the same time or place, allowing discussions and reflections to take place more gradually and completely over time as well as across arbitrary distances.

**The Perspectives Mechanism**

The technology currently implemented in WEBGUIDE and envisioned for the POW! Perspectives server supports the construction of knowledge in personal, group and comparison perspectives. The server allows users to define a network of interconnected perspectives which inherit content from each other so that my personal perspective automatically contains ideas that my team has already agreed on in its group perspective and a comparison perspective automatically contains ideas from my personal perspective and from those of selected colleagues. New perspectives can be added by users on the fly.

The perspectives server keeps track of all the relations among perspectives and ideas of different people. It prepares content views transparently so that client interfaces can navigate the perspectives and ideas intuitively. Users can articulate, reflect upon, modify, compare and negotiate ideas in the
shared, evolving collaboration space without worrying about the underlying structure of the perspectives.

While much collaboration software could benefit from a perspectives mechanism, no other system has as versatile a perspectives mechanism as WEBGUIDE. Some systems have simple mechanisms, perhaps allowing several personal perspectives and one group perspective _ fixed in structure and lacking inheritance of content. Most collaboration systems have no such facility. Other researchers are interested in incorporating WEBGUIDE_s perspective mechanism once it is available as an open source server.

**Approach to Domain Knowledge:**

The PI and Graduate Student will attend corporate meetings of the training network to which the Collaborating Company belongs and will participate (as facilitators and observers) in corporate meetings and training sessions conducted by the Collaborating Company. These sessions will be preceded and followed by debriefing sessions with the Business Representative. Project staff will meet regularly with the Collaborating Company to collaboratively develop requirements for corporate applications of the software.

**Approach to Software Architecture:**

The PI has already acquired a Linux webserver with dual Pentium processors, fast database access and a high-speed Internet connection for use in the POW! Project. A Java application _ the POW! perspectives server _ will run on this webserver and will access a mySQL database. A light-weight client will run in the browser of a user running on any platform (Mac, PC, Unix, Linux). The client and server will communicate using CGI calls and XML data formats, allowing secure communication through firewalls. The calls will be optimized to enhance cross-Internet performance and maximize client display speed.

**Approach to Intellectual Property:**

The separation of applications into an Open Source Software (OSS) server and proprietary client will allow the Collaborating Company and other Colorado businesses to develop software for their own applications quickly
and flexibly, while making use of the computationally complex perspectives mechanism seamlessly. This takes advantage of the benefits of both the traditional economic model and the new open source approach: development of the general mechanism can be shared while specific applications can provide economic competitive advantage. The details of the POW! license will be negotiated with the University and the Collaborating Company and will be designed to foster these complementary advantages.

**Approach to Application Clients:**

Internet technology is evolving rapidly. Compatibility with hardware and software in use at different sites is a major problem. The POW! perspectives server will run on a webserver, such as ours at CU, and does not need to be compatible with a variety of user systems. Some application client developers may want to take advantage of the latest versions of Java while others may prefer to maintain compatibility with older versions of HTML. The architecture developed by this Project will allow developers to create client interfaces using HTML forms, Perl scripts, Java applets and other technologies (such as XSL stylesheets). The Project will develop, document and post three sample clients demonstrating how to program client software using these different technologies and still taking advantage of the perspectives mechanism.

**Resources**

The University and the Collaborating Company already have adequate office space, computers, commercial software and networking to support this project. The University will only need funds for computer support. The PI will contribute technical expertise and the Collaborating Company will contribute expertise in supporting corporate meetings and trainings.

**Evaluation Plan**

The POW! Project will be evaluated by the production of the following deliverables:

- a POW! perspectives server in Java with a clear API and XML data exchange.
• negotiation of a license and the release of the POW! server under this license on a website with appropriate documentation to support collaborative development.

• sample clients for the POW! server illustrating the use of Java, HTML and Perl technologies.

• a requirements document for corporate client applications.

**Follow-on Funding Plan**

The POW! Project will provide a foundation for future work along two dimensions:

• The PI will raise over $100,000 in federal funds to continue work by him and the graduate student in educational applications using the POW! server.

• The Collaborating Company will raise funds internally and/or through investors to continue the development and marketing of software clients for corporate applications in collaborative distributed decision-making and training.

**References**


Perspectives On the Web (POW!)

Proposal to Intel Corporation

RESEARCH OBJECTIVE

Summary: This project will develop effective systems to represent personal and group perspectives on the Web. Computational support for perspectives can significantly enhance the potential power and usability of Web-based collaboration systems for groups of learners or workers. This can help break down the barriers to widespread participation in computer-mediated collaboration for the development of ideas and allow computer-based knowledge-building communities to proliferate around the world.

Problem: Collaboration is highly recommended for education and is a fact of life in contemporary work. But collaboration is very difficult, particularly in situations that require the building of new knowledge by the collaborating group. Collaboration requires a complex interplay between individual and group tasks, that are hard to keep track of and to manage. Computer support of collaborative learning (CSCL) and of cooperative work (CSCW) promise important advantages, such as: independence from people having to be at the same place at the same time, persistence of ideas, tools for analysis and reflection, access by extended virtual communities. However, current systems like threaded discussion forums suffer from serious weaknesses in practice. Discussions tend to be short-lived, inspire low participation, result in divergence of ideas, and remain superficial.

Technology and Usage Issues: The proposed POW! Project addresses the barriers to technology adoption in an important area of information management: an area identified in this proposal as "collaborative knowledge-building environments." These systems promise to provide effective environments for education and knowledge work. They represent a new usage paradigm associated with the facilitation of effective online
communities. The proposed system design approach is responsive to the social implications of the technology in typical use contexts.

Software environments to support collaborative knowledge building need to incorporate a well-designed set of computational capabilities and they need to be used within a carefully orchestrated set of social usage conditions. The POW! Project will design, implement, field and assess a technology of computational perspectives within a knowledge-building environment. This technology, when configured and used within the proper social conditions, will represent a set of personal and group perspectives of the user community. Users collaborating in the environment will be able to define their own perspectives on common topics, exchange those perspectives, and negotiate shared group perspectives.

A computational perspective is a restricted view into a larger information space. It is defined by an individual or group to capture content of interest to them and to modify that content in their own view. The POW! system will provide support for defining and elaborating alternative perspectives on ideas. Relationships can be established among the perspectives, so that they benefit from each other without disturbing each other. The software supports an arbitrarily complex graph of perspectives for groups, multiple layers of sub-groups, individuals and even alternative perspectives for a given individual. There is a mechanism of content inheritance whereby perspectives incorporate the content of higher level group perspectives, so that there is sharing of ideas and each perspective does not have to start from scratch. Ideas can be negotiated so that they migrate from personal opinions to shared understanding and group knowledge. The perspectives are supported within a Web-based environment for knowledge construction and collaboration that provides a range of useful functionality.

**Research Hypothesis:** Using computational personal and group perspectives to structure the display of shared information spaces can facilitate collaborative knowledge-building processes.

**Verification:** Analysis of the usage of software prototypes using Perspectives On the Web will evaluate the increased effectiveness of collaboration in knowledge-building processes. Studies of threaded discussion systems lacking computational perspectives (dePaula, 1998; Guzdial & Turns, forthcoming; Hewitt et al., 1998) have demonstrated their weaknesses with the following measures:

- short threads: a tendency for discussions to die quickly,
- low participation: a lack of motivation to collaborate,
• few cross-references: little convergence of ideas,
• superficial content: minimal depth of investigation.

The POW! Project will field test software prototypes in a variety of contexts during a three year period:
• college seminar,
• corporate training,
• academic research group,
• corporate work group.

The resulting discussion will be analyzed using a variety of methods, including:
• structural analysis: quantitative statistics on thread lengths, participation levels, etc.,
• discourse analysis: qualitative analysis of collaborative exchanges,
• content analysis: coding of contributions and implicit cross-references,
• survey analysis: interviews and surveys of user experiences.

The results of the four analyses under the four conditions will be compared with previous studies without perspective support to verify (or falsify) hypothesized improvements of measures the four measures.

Potential Impact: Organizing contributions to knowledge-building processes on the basis of personal and group perspectives — assuming it proves to be effective — could contribute to a breakthrough in computer support of collaboration. Collaboration provides a powerful — virtually necessary — approach to learning and working in the modern and future world. Breakthroughs in computer support of collaboration could result in widespread use of computer mediation at all levels of schooling, employment and community interaction. The impact will be particularly dramatic for effective distance education, interactive organizational memories and engaging virtual communities.

RELATIONSHIP TO OTHER RESEARCH & PRACTICE

Related Work: The importance of the concept of perspectives has long been recognized in the philosophical and theoretical literature (Boland &
Tenkasi, 1995; Gadamer, 1960/1988; Heidegger, 1927/1996). The PI’s dissertation work defined a perspectives mechanism within a stand-alone hypermedia system for designers (Stahl, 1993a). In the following seven years, he developed this approach further, eventually implementing perspectives on the Web and the WEBGUIDE environment for collaborative learning (Stahl, 1995; 1999c; 2000; Stahl & Herrmann, 1999). Other hypermedia systems using perspectives have stalled after generating initial enthusiasm (Boborow & Goldstein, 1980; McCall et al., 1990; Mittal et al., 1986; Nelson, 1981). Educational knowledge-building environments represent an important research area currently, but none of them incorporate perspectives support for groups and individuals (Guzdial & Turns, forthcoming; Hendricksen, 1999; Pea, 1993; Scardamalia & Bereiter, 1996; Slotta & Linn, forthcoming).

**Uniqueness of This Work:** Commercially available threaded discussion systems are now abundant; however, they suffer from serious weaknesses and need to be incorporated within richer computational contexts. Research prototypes of collaborative knowledge-building environments are part of a vigorous sub-field within CSCL and CSCW. Some of them use the term _perspectives_ in the sense of alternative displays (e.g., textual versus graphical) of the same data. Others provide personal views of just one user’s own contributions. But none represent relationships among personal and group perspectives through an arbitrarily complex graph of perspectives (Stahl, 1999a). The POW! Project takes the relation of personal and group perspectives as a key structure of collaboration, tries to provide a computationally powerful and intuitively useful representation of that structure, and assesses its effectiveness.

**Use of Related Work:** The POW! Project is firmly situated within the active research sub-field of Collaborative Knowledge-Building Environments (CoKnowBEs). The PI is currently leading a research seminar on this sub-field and organizing a related workshop at CSCL ’99. In addition, he is the PI on a CILT seed grant from SRI International to facilitate collaboration within this research community. The goal is to foster interoperability among related prototypes and systems so that the POW! Project prototype can incorporate data and functionality from other developers and they can ultimately incorporate the functionality of perspectives.

While the POW! Project has a specific, unique focus around the technology of computational perspectives, it is an integral part of a much more general effort to advance a broad class of collaborative knowledge-building environments for learning and working. The technological and (even more
so) the social research issues that must be solved for these systems to realize their tremendous potential are far too complex to be solved by an isolated research effort. Building upon the POW! Project, the PI and his colleagues are compiling a shared knowledge base related to CoKnowBEs, to promote data interoperability among CoKnowBEs, to eventually build interchangeable CoKnowBE components and to generally increase the level of collaboration among developers and researchers concerned about this class of systems.

**Long-Term Research:** The idea of computational perspectives was explored at Xerox PARC 20 years ago (Boborow & Goldstein, 1980) and the idea of collaborative knowledge-building environments has been pursued for 10 years (Scardamalia & Bereiter, 1996). Yet, there are still important open research questions _ both technical and social (Kling, 1999)_. The Web has been wildly successful with applications that are structurally very simple, trivial to use and relatively easy to understand. Collaborative knowledge building is inherently complex, requires subtle interpersonal skills and is hard to comprehend. While it seems inevitable that the world move rapidly toward ubiquitous networked collaboration, this will not happen as automatically as Web browsing, email correspondence or e-commerce. First, researchers will have to develop well-designed sets of functionality to support the structures and processes of collaboration and carefully orchestrated social procedures will have to be put into place. Significant progress in this direction can be expected in the next decade if efforts are made to address these issues systematically.

**Research Strengths:** The PI has been working on software incorporating perspectives for almost a decade. His WEBGUIDE software has already been developed in an initial prototype and fielded in two use situations. He works within a research center that is well established and has been exploring the design of high-functionality software and computational support for learning and work for over 15 years. The PI is in working contact with researchers in the US, Canada, the UK and Germany engaged in related work. The Project will have access to appropriate test sites and will leverage supplementary funding for colleagues and equipment.

Complementary Work: The PI is currently involved in the following complementary projects:

**Organizational Memory and Organizational Learning**_ in the Center for LifeLong Learning & Design explores computer-based group memories and the use of conferencing systems.
Articulate Learners in the Institute of Cognitive Science explores the use of latent semantic indexing in educational technology, with an emphasis on rigorous testing and evaluation.

Symbiotic Computing (proposed) in the Department of Computer Science explores pervasive networking of servers, desktop computers, laptops and very small distributed and embedded devices to support collaboration using shared information spaces.

Data Interoperability in CoKnowBEs sponsored by CILT at SRI explores collaboration among researchers in North America, including the definition of XML standards for exchange of data among knowledge-building environments.

RESEARCH POTENTIAL IMPACT

Expected Results: In addition to conference presentations, journal papers and a monograph reporting on the conceptual framework, software prototypes, field testing and effectiveness assessment, the following concrete products are expected:

• A PowServer Java middleware application that accesses the shared database, computes displays in specified perspectives and returns a tree of data to a client.

• A WEBGUIDE Java client applet that provides a user interface in a Web browser to support collaboration.

• Analysis of field tests of the software in a variety of usage contexts.

• A set of XSL templates for displaying perspective data stored in a canonical format as an XML file.

Applications of Results: The PowServer and the XSL templates will be released as Open Source so that they can be used by anyone who develops a compatible client application. The WEBGUIDE client will be documented as an example Java client, but other clients based on HTML, Perl and CGI scripts could also be used. The PowServer API will be documented. The analyses of field tests will suggest how the perspectives software can be used effectively.

Benefits to the Computing Industry: The proliferation of more sophisticated, high-functionality applications on the Web will open up the Internet to a greater variety of uses and users. The potential of networked
computation and collaboration will be further realized and this will drive the spread of ubiquitous computational devices.

**Required Advances:** Networking bandwidth (multi-megabit wireless access), computational power, memory capacity, device portability (size, weight, power consumption), multi-modal input and display interfaces (speech, large flat panel), seamless global roaming, acceptance of infrastructure of social practices. The technological advances are currently proceeding at a pace to make pervasive networked collaboration practical within the coming decade if the software design and social practice issues become well enough understood.

**RESEARCH PLAN**

The POW! Project is already underway. The basic perspectives technology is already designed (Stahl & Herrmann, 1999). WEBGUIDE, a preliminary prototype has already been developed and field tested in two educational contexts, a middle school environmental science class and a graduate cognitive science class (Stahl, 1999c). An informal analysis has determined weaknesses that must be addressed in the next round of work (Stahl, 1999b).

Experience to date indicates that there is no clear path to solving the complex social, educational and technical problems raised by this Project. A cyclical approach of thoughtful trial is needed. Each year, the proposed project will cycle through the phases of (a) theoretical reflection, (b) system prototyping, (c) practical application and (d) functional assessment. In Spring semester, the PI will conduct an interdisciplinary graduate seminar on the cognitive theory of knowledge-building environments. In the summer, the PI and colleagues will develop or revise prototype network-based software with functionality suggested by the theory and results of previous trials. In the Fall, the prototypes will be used in educational and work contexts. Assessments of the use of the prototypes will be conducted, published and incorporated into the next cycle.

Year 2000 Objectives: Assessment of the first WEBGUIDE prototype revealed the following weaknesses:

- Response time was too slow, especially on slow middle school computers at remote locations or for colleagues in Europe.
- Graph structure of perspectives and of discussion notes was not clear to users.
• Users were confused about which perspectives to use for discussion and for development of personal ideas.

• Despite functionality to permit and encourage convergence of ideas (topic structure, multiple linking, comparison perspectives), discussions diverged and died.

• Data from different versions of WEBGUIDE and from other similar software is incompatible and cannot easily be merged or compared.

The next version _ WEBGUIDE 2000 _ therefore has the following year one objectives:

• Separation of WEBGUIDE into a PowServer running on a fast Web server to do intensive computations of perspective displays and a client applet to handle navigation within a perspective and to provide user interface services. Optimize Internet communication through the client/server API.

• Provide an optional graphical view of the perspective and note structure, with capabilities to add, delete, edit and link notes together. Increase the ability of the user interface to display large discussions.

• Revise discussion support and add support for private notes.

• Add negotiation support, including summarization, voting, debating and promotion to group perspectives.

• Provide for import and export of data to a standard XML format and for display using XSL.

Deliveryables:

• A PowServer Java middleware application that accesses the shared database, computes displays in specified perspectives and returns a tree of data to a client. (Probably a revised version each year.)

• A WEBGUIDE Java client Web application that provides a user interface in a browser to support collaboration. (Probably a revised version each year.)

• Annual analysis of field tests of the software in a variety of usage situations.

• A set of XSL templates for displaying perspective data stored in a canonical format as an XML file.
• Quarterly reports.
• Copies of works published as a result of this research.

**Personnel:** Only the PI will be supported from this grant. Related grants will support a graduate research assistant and an undergraduate research apprentice. A variety of colleagues, students and researchers at other institutions will collaborate with this Project indirectly.

**Milestones:**
• Release of a PowServer application running on a Web server.
• Release of a WEBGUIDE client applet available on the Web.
• Field trial in a classroom, research project group, corporate training or business network.
• Analysis of field trial results.

**Tradeoffs:** A high-functionality Web-based application must always trade-off functionality versus ease of use and understandability. Scarce screen real-estate, response delays and limitations of different browsers and platforms further complicate software design. The proper balance requires experimentation and field trials to see what is most important and which problems can be tolerated.

**Risk Management:** The hardest problem will be to match customized versions of WEBGUIDE and specially configured databases to field trial opportunities. Good opportunities come with fixed timetables, idiosyncratic priorities and a life of their own. It can take a lot of work to give the field trials a reasonable chance for success. The POW! Project will take advantage of a variety of kinds of trial opportunities, including some that the PI has considerable control over (e.g., his own seminars). The Project will develop a small number of high quality trial opportunities (one or two a year) rather than trying to take on more trials than it can adequately support and carefully assess.

**Technology Exchange:** The PowServer technology will be available to Intel engineers and they will have access to the databases of field trials. It is possible that in Year 3 a work group within Intel will want to try collaborating using the technology.
The Research CyberStudio: Supporting Researchers as LifeLong Learners

The Research CYBERSTUDIO Project addresses the problem of training and supporting learners at the most advanced end of the educational system to be skilled interdisciplinary researchers. It targets graduate students who have completed extensive classroom study within a discipline but who could benefit from practical research experience within a supportive context. The project’s theoretical perspective of lifelong learning postulates that people in knowledge-intensive endeavors need to be continuously developing skills and constructing knowledge, and that this can be facilitated by information delivery technologies within supportive collaborative contexts. Accordingly, the proposed project approaches its goals from a technological and organizational approach: creating structured communities of learners (research studios), and providing adaptable computer-based support (the CYBERSTUDIO) for these communities. The technical approach builds on innovative software prototypes by the PI and collaborators; the organizational approach leverages substantial local opportunities. While the project is designed to assist novice researchers, it will develop computer software useful to interdisciplinary research communities generally.

THE PROBLEM OF TRAINING AND SUPPORTING RESEARCHERS

Research is an important aspect of contemporary universities like the University of Colorado (CU). Increasingly, much of this research is taking on an interdisciplinary character, spawning special groups like the Institute of Cognitive Science (ICS) and the Center for LifeLong Learning and Design (L3D) at CU.

Despite a broad national effort to reform education from kindergarten through college and a significant attempt to develop computer support for education, little has been done to address the educational and computer needs of the most advanced students. The transition from an educated domain specialist to a skilled researcher is a lengthy and haphazard process, largely because the student is given little systematic support (Denning 1992). It is assumed that once students have completed their graduate course
work they are capable of pursuing dissertation and post-doc research with minimal pedagogical support. However, experience within ICS and L3D shows that fledgling researchers need to continue developing their skills in reading, writing, and mathematics just like students at any level. As they become involved in investigating problems that spill outside the discipline of their academic training, they need to learn to read broader professional literature, to prepare journal articles or conference presentations, and to master new methodologies (statistical evaluation, experimental design, computer modeling, discourse analysis, etc.). Whereas most professionals have specialized productivity software at their command, interdisciplinary researchers lack such tools.

The Research CYBERSTUDIO (RCS) Project adapts current constructivist educational theories to the problem of training researchers. In particular, the following pedagogical principles underlie the Project’s approach:

1. Knowledge is constructed within communities of learners (Scardamalia & Bereiter 1994).
2. The approach of a design studio provides an effective setting for learning (Schön 1987, 1983).
3. Learning takes place through a person’s increased participation in a community of practice (Lave & Wenger 1991).
4. Individual understanding can be fostered by appropriate computer-based systems (Papert 1993, 1980).

Based on these principles, the Project conceptualizes the problem of training novice researchers in the following terms:

1. The learner is viewed as a newcomer within a research community, as an apprentice who needs thoughtful mentoring.
2. Collaborative research activities are organized into a “research studio” structure in which individual and group projects are conducted and critiqued.
3. Learners are assisted in gradually participating more and more in their research community to acquire the tacit skills of their profession.
4. Special software acts as a “CYBERSTUDIO” in which community members communicate and contribute, work and learn.

The uniqueness of the RCS Project lies in the creation of computer support to promote lifelong learning and to manage organizational knowledge within
a research community. The CYBERSTUDIO software will therefore be described first. Then its usage by the community will be discussed.

THE CyberStudio SOFTWARE

A central hypothesis of this project is that computer support can play an important role in developing interdisciplinary research skills as well as in accomplishing the research itself. The challenge of the project is to create new software adequate to the attainment of this potential.

The lack of computer support to deliver information on an as-needed basis is endemic to interdisciplinary research in general, not merely to novice researchers. The theories of situated learning and knowledge construction suggest that there is a significant untapped potential of computer support for building communities of learners and for capturing group memories to inform newcomers. Such knowledge sharing software could be particularly helpful in the interdisciplinary context, where community members speak different technical languages.

Project participants will design, prototype, explore, and evaluate CYBERSTUDIO’s software to support learning, communication, and work within interdisciplinary research groups. This software will be designed to meet the information and collaboration needs of researchers, especially novices. The success of the community-of-learners approach requires a high level of communication and organization; CYBERSTUDIO will provide a medium in which this can take place. The software will also identify and deliver relevant ideas from the extensive and growing writings of the group and related published literature, allowing people to share ideas across time and space (Stahl et al. 1995a, 1995b).

Here is an illustration of how the Internet-based CYBERSTUDIO software can be used: Suppose that a graduate student drafts a thesis proposal for software to categorize the content of Internet sites by reading level. The proposal text is analyzed within CYBERSTUDIO. The software delivers a list of web links pointing to the most directly related excerpts from cognitive science papers, key terms in an interactive, multidisciplinary glossary, specific entries in threaded discussions within the research groups, email messages on the topic, and contact information for people in various disciplines who have done relevant work. The student can then review and respond to any of this information. For instance, the student might compile a set of notes with hypertext links to several of the retrieved sources, make annotations to the sources (for future users to read), send messages to referenced people. Then the student can revise the proposal draft and
resubmit it to CYBERSTUDIO to obtain a refined list of relevant information. All of this is done within the CYBERSTUDIO system.

While the CYBERSTUDIO repository of information is primarily directed internally to the research group, it also includes external links to web sites globally and it allows outsiders to view many materials in the network. Thus, it provides a medium of communication and documentation within a local research community while participating in the broader discourse of the World Wide Web.

The CYBERSTUDIO software system envisioned to support working and learning by interdisciplinary researchers unifies three technologies that the PI and collaborators at L3D and ICS have been exploring for many years:

1. Domain-oriented design environments (DODEs).
2. Dynamic web sites (DynaSites).
3. Latent semantic analysis (LSA).

These technologies will be integrated into CYBERSTUDIO network of research information services. CYBERSTUDIO captures knowledge as it is constructed within a research group and delivers items from this organizational memory when they are relevant to the new research of individuals, particularly newcomers.

1) Domain-oriented design environments. A DODE is a software application within which a professional conducts work. As the work progresses, the software responds by delivering domain-specific or community-historic information stored in its knowledge base that is relevant to informing the current state of the work (Fischer et al. 1993a, 1993b). By integrating working and information delivery, DODEs support lifelong learning or learning-on-demand.

2) Dynamic web sites. DynaSites are web sites that provide an interactive interface to a database of information shared by a group (Stahl 1997a). Based on intranet technology, they transform the World Wide Web from a generic broadcast medium to a group memory that allows collaborators to share their knowledge asynchronously. For instance, people in a research community can use their web browsers to find past discussions of ideas, glossaries of terms, and papers published by other members; as they generate new ideas, concepts, and essays using these resources, the new knowledge is added to the group memory interactively.

3) Latent semantic analysis. LSA is an automated technique for analyzing the semantic relations within a large corpus of text (Landauer & Dumais
LSA can compare documents and rate the similarity of their technical content. When used properly, it can be effective for such tasks as evaluating the knowledge content of essays. Thus, it can be used by software to judge which of several student essays is most similar to a target essay. Experiments have shown that LSA is approximately as reliable as people in grading SAT essays and in selecting readings that are most appropriate for a given reader based on an essay by that reader (Wolfe et al. 1997).

The CYBERSTUDIO software developed by this project will incorporate information sources relevant to the research of the interdisciplinary groups involved (e.g., L3D and ICS). This includes both archival materials (published papers, technical reports, dissertations, seminar presentations) and process artifacts (on-going threaded discussions, email exchanges, meeting schedules or minutes, evolving glossaries of technical terms, annotated bibliographies, member information, etc.).

The project involves the development of techniques for capturing, structuring, evolving, retrieving, and presenting the information in CYBERSTUDIO. These techniques will include LSA applications (including the semi-automated production of a glossary of interdisciplinary technical terms in a corpus linked to key document excerpts), group perspectives (Stahl 1997a, 1995a, 1993a, 1993b), and a visual end-user language for querying and navigating the information base (Stahl 1993b, 1992a, 1991). Development of these techniques will be staged during the RCS Project period, with the glossary developed in Year I, perspectives in Year II, and the end-user language in Year III.

THE ORGANIZATIONAL APPROACH

The research studio is an approach to training through self-directed hands-on experience. Architecture students, for instance, spend a lot of their class time in studio classes, where they work on individual or group projects and receive critiques from peers and experienced designers. The project’s research studios build on this model. Novice researchers will pursue their own dissertation research or participate in funded research within a community of learners, including both more and less experienced researchers. In addition to interacting informally and making formal presentations, people will share and co-construct ideas in settings such as reading groups, project meetings, and on-line discussion threads. Much of the communication associated with research studio activities will take place within the CYBERSTUDIO system and will be captured by it. Then, future newcomers can review the materials to learn relevant aspects of the group’s intellectual history.
The project will investigate effective ways of structuring interdisciplinary research groups. This includes issues of physical office arrangements, meeting procedures, study groups, communication channels, and decision making. Like most professional workers, graduate students have too much to do; they must resolve conflicts of course work vs. research, individual projects vs. group efforts, meeting vs. working, learning vs. producing. Some of these conflicts can be ameliorated via institutional solutions such as adjusting requirements and reward structures. The framework of a research studio will be explored as a way of integrating research practice into the academic reward system, so involvement in group activities does not detract from personal achievement.

Apprenticeship or mentoring is important to the studio model of situated learning. A more experienced person provides systematic guidance or facilitation of student self-directed learning, and the student learns by working alongside old-timers. The mentoring relationship—just like the research studio—must be institutionally recognized in order to be effective. The project will investigate how this can be accomplished. It will start by formalizing apprenticeship relationships in the sense that they will be explicitly recognized within the group. Both the mentor and the apprentice will receive recognition for their work together. An on-going dialog concerning benefits and problems of apprenticeship will evaluate this approach.

Because the RCS project is itself an interdisciplinary research effort involving reflective practitioners, the participants in the investigation will have the task of evaluating their own learning. They will incorporate evaluation methodologies from multiple disciplines (educational evaluation, psychological controlled studies, software engineering debugging, user testing, etc.). Assessment will itself be a topic of research—how to evaluate support for lifelong learning and interdisciplinary research in naturalistic settings.

THE LOCAL CONTEXT

This project addresses the problem of training and supporting researchers from within an exceedingly rich context of growing interdisciplinary research at the University of Colorado (CU). It will take advantage of considerable independent resources from federal, foundation, and university sources and focus them on the needs of interdisciplinary researchers.

Project level. The PI is currently directly involved in two interdisciplinary research projects: an effort to develop computer-based organizational
memories and one to develop educational software. The first is sponsored by the Center for LifeLong Learning and Design (L3D) and the second jointly by L3D and the Institute of Cognitive Science (ICS). The organizational memory project integrates ideas from learning theory, anthropology, and organizational theory as well as various aspects of computer science and particular application domains. The educational software project involves issues of psychology and linguistics as well as computer science and education. During Year I, the RCS Project will encompass the teams of graduate students, post-docs, and visiting researchers working on these two projects.

Center level. ICS is an interdisciplinary institute by the nature of cognitive science; it expects to become an accredited interdisciplinary degree program in the next year. Within ICS, there is an active research group exploring latent semantic analysis (LSA). LSA is a statistical text analysis method with promising applications to practical problems in educational software as well as theoretical implications within cognitive science. The LSA research group includes cognitive psychologists, computational linguists, and computer scientists. L3D is a center under both ICS and CU’s Department of Computer Science, with strong involvements in education and environmental design. It encompasses projects developing conceptual frameworks and prototype software for applications in a variety of domains. L3D and ICS members teach undergraduate and graduate courses in computer science, design, and cognitive science. Students in the research groups and courses within L3D and ICS will provide the focus for Year II of the RCS Project.

University level. The CU administration is promoting the notion of a “total learning environment.” As part of this commitment, L3D is establishing a broader interdisciplinary initiative across many departments of CU—the Center for Interdisciplinary Research on LifeLong Learning (CIRLL)—likely to be funded as an NSF center next year. CIRLL will directly support seven graduate research assistants and four post-docs, as well as coordinating the work of many more novice and experienced researchers across campus. In addition, L3D has a growing network of industrial partners; students intern at the companies and company employees spend time at L3D’s research labs. In Year III, the RCS Project will expand to include novice interdisciplinary researchers in CIRLL and among the industrial interns.

Broader impact. It is anticipated that the lessons learned in the RCS project—pedagogical approaches, organizational supports, and computer software designs—will be disseminated beyond CU through research
contacts at key centers like the Cognitive Studies of Interdisciplinary Communication program in the National Institute for Science Education at the University of Wisconsin, as well as through academic publications. Within CU the potential for dissemination is unlimited, with CU’s focus on “total learning,” its efforts to promote Internet support for teaching, its interest in distance learning, and its standing as a major research center.

CONCLUSION

The Research CYBERSTUDIO Project will explore organizational, pedagogical, and technological approaches to train advanced graduate students to be skilled interdisciplinary researchers. It will develop CYBERSTUDIO software to support the work of research groups. Gradually expanding its scope, the project will take advantage of a lively and growing community of interdisciplinary research at the University of Colorado. The approaches and software developed will be thoroughly evaluated, clearly documented, and broadly disseminated.

References
