Optimizing Feedback for Eliciting Pedagogically Valuable Explanation in Collaborative Problem Solving

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 empirical work on which we build in the areas of computer supported collaborative learning and tutorial dialogue systems, and with a practical, concrete plan to demonstrate impact within the three years of the proposed work.

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. In support of this goal, we propose advances to the language technology field in the area of semi-supervised approaches to multi-class classification and applied dialogue agent technology.

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.

Optimizing Feedback for Eliciting Pedagogically Valuable Explanation in Collaborative Problem Solving

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 promises to address this problem by providing free or inexpensive 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 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. Through PDP, support for productive patterns of collaborative learning has been shown through quasi-experimental studies to be directly linked with increased success in math for minority and low income students. Making what would normally be a staff intensive program available ubiquitously at a dramatically reduced expense would be an enormous payoff. 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, Mulis, 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. 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 empirical work on which we build in the areas of computer supported collaborative learning and tutorial dialogue systems, and with a practical, concrete plan to demonstrate impact within the three years of the proposed work.

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. This will be accomplished through close collaboration among CMU, Math Forum and VMT researchers. As a starting place we begin by integrating research findings and infrastructure from our prior work in the areas of computer supported collaborative learning and tutorial dialogue systems. We describe this integrated foundation in Section 2. In Section 3, we then describe a series of proposed intervention studies to be conducted in a lab setting in which we will further explore the design space for adaptive collaboration support meant to stimulate and shape social and educational interaction among learners. Building upon the technological foundation provided by our prior work, the goal of these interventions is to improve students' attitudes toward their mathematical learning and the on-line learning environment overall, to increase engagement with the course material, and to improve learning outcomes. In order to move these interventions from the lab into the classroom, and ultimately into "the wild", we must build on and extend our existing technological infrastructure. Thus, in Section 4 we describe how we propose to extend language technologies, including tutorial dialogue agents, machine learning technology, and other technologies that allow agents to automatically recognize opportunities for intervention and allow us to easily construct human-machine dialogs to form these interventions. Section 5 describes our plans to evaluate our fully implemented interventions in the classroom as well as in the naturalistic, on-line VMT environment. Sections 6 through 8 describe key partnerships essential for the success of the proposed work, results from our prior NSF support, and a timeline for the proposed work respectively.

2. Building a Foundation by Integrating our Prior Work

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. Researchers have examined the mechanisms by which human tutors are so successful at teaching and motivating children as a model of successful education (Bloom, 1984; Cohen, Kulik & Kulik, 1982). Unfortunately, it is not practical to provide every student with a human tutor. While there is a shortage in terms of the resources to provide each student with their own tutor, there is no lack of children in need. 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; Gweon et al., submitted), 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), and we can use this automatic analysis to trigger interventions that support students in helping each other learn together (Gweon et al., 2006), we could move towards a solution to our nation's educational problems in a cost effective, practical manner. In this section we describe how we integrate elements from our previous work into a technical foundation as well as a foundational instructional approach that we build on and extend in our proposed work.

Selecting appropriate materials to stimulate productive collaborative conversations is essential to 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.

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 4.1) 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, adaptive 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.

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 automating collaborative learning process analysis (Donmez et al., 2005), other language technologies research related to text classification (Rosé et al., 2003; Rosé et al., 2005; Donmez et al., submitted), 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 work on design and evaluation of adaptive collaborative learning support (Gweon et al., 2006) and investigations of group composition and gender effects in collaborative learning (Gweon et al., 2005; Gweon et al., submitted).

Our instructional approach is modeled after constructivist principles of classroom discourse, such as those advocated in (Chapin, O'Connor, & Anderson, 2003). 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 (Rummel et al., 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., submitted).

Meloth and Deering (1999) present evidence of the importance of the teacher's role in supporting effective collaborative learning. The teacher's input is essential for keeping group discussions moving in a productive direction. The teacher provides key insights and models a productive learning process. It is this type of support we seek to capture in the adaptive collaboration support we propose to develop and optimize in this proposal, building upon our prior work along these lines (Gweon et al., 2006).

> **Eddie:** Well, i don't think it matters what order the numbers are in. You still get the same answer. But a half of three quarters and three quarters times a half seem like they could be talking about different things.
> **Teacher:** Rebecca, do you agree or disagree with what Eddie is saying?
> **Rebecca:** Well, I agree that it doesn't matter which number is first, because they both give you the same answer. But I don't get what Eddie means about them saying different things.
> **Teacher:** Eddie, would you explain what you mean?
> **Eddie:** Well, I just think that like different orderings can refer to a different type of process going on.
> **Tiffany:** But you still have the same result, so they are the same!
> **Teacher:** OK, so we have two different ideas here to talk about. Eddie says the order does matter, because the two orders can be used to describe different situations. So Tiffany, are you saying that the two orders can't be used to describe two different situations?
> **Tiffany:** No, …

> **Teacher:** OK, we need to think more about Eddie's statement. I Eddie's statement he is saying order matters because the process is different, not because of whether the answer is the same or not.

Note that the teacher in this example plays a very active role in the conversation in directing students to explain things that are unclear, interpret other student's contributions, and address other student's contributions. The teacher uses questions strategically to stimulate cognitive conflict and reflection, drawing student attention to issues that the teacher believes would be productive for students to address, and elevating the level of *transactivity* in the dialogue, which is the extent to which students directly address the contributions of others in collaborative discourse (Weinberger & Fischer, in press; Teasley, 1997). But the teacher is not engaging in Socratic tutoring practices per se, where the teacher would be leading students down a predetermined directed line of reasoning, as in (Rosé et al., 2001). Instead teachers plan activities that they think will stimulate a variety of ideas and issues. Then they behave opportunistically, directing their feedback in ways that intensify discussion of these topics when the students bring them up. In this way, the learner's path is determined through a negotiation in a sense between student initiative and teacher initiative.

The patterns of classroom discourse that we see advocated are very similar in spirit to the Negotiable Problem Solving Goals paradigm that underlies the design of our CycleTalk tutorial dialogue system, which has been successful in classroom evaluations with college aged students (Rosé et al., 2005; Kumar et al., 2006). Thus, we situate our instructional approach along what we refer to as the Exploratory Learning Continuum, which has provided a conceptual framework for our previous work on guided exploratory learning. Exploratory learning has typically been associated with "high level goals" such as "survive in this simulation environment" or under specified goals such as "find all implications that can be drawn from these premises". In contrast, non-exploratory learning has been associated with "low level goals" or "fully specified goals" such as "solve this equation" or "verify whether this implication is true". We argue that all learning is exploratory, and alternative learning tasks or learning environments can be placed along a continuum (See Figure 1).

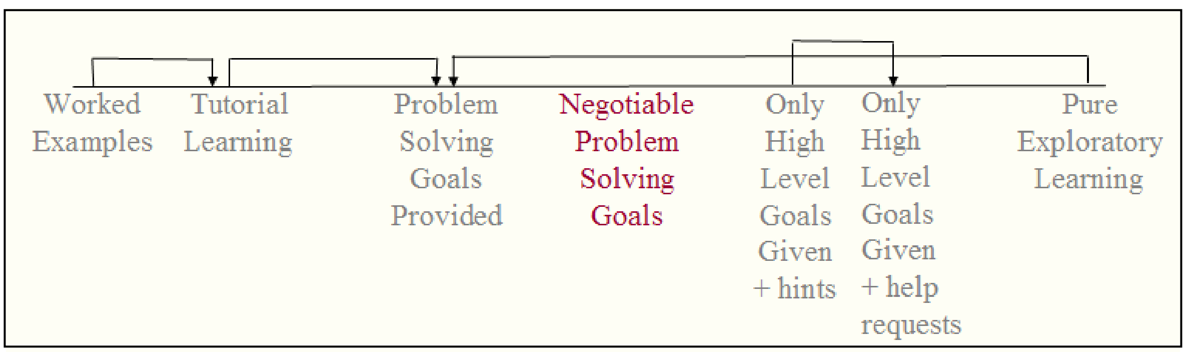

Figure 1 The Exploratory Learning Continuum: Arrows connect approaches that have been experimentally compared in published works. The arrow points from the less effective approach towards the approach shown to be more effective.

Many state-of-the-art tutoring systems fall into the problem solving category where problem solving goals are dictated. It is no coincidence since published investigations along the Exploratory Learning Continuum have typically shown this place on the continuum to be particularly effective. For example, Charnay & Reder (1986) compare Worked Examples, Tutorials, Problem Solving, and Pure Exploration. Worked examples mixed with problem solving was the best combination, consistent with other similar published results Charnay & Reder (1986). Along similar lines, Klahr & Nigam (2004) have shown in an empirical investigation of children learning the scientific method that tutorial based learning mixed with problem solving is more efficient than pure exploratory learning. Other work has explored a part of the continuum in between problem solving and pure exploratory learning. In the light of a series of previous results showing the benefits of guided exploration over pure exploration (e.g., Leutner, 1993), the Smithtown work (e.g., Shute & Glaser, 1990) and the Computer-Based Simulation Games work (Mayer, 2004) involve guidance provided by high level goals such as learning about a model or survival in a simulation environment. Leutner (1993) demonstrates the importance of students with prior domain instruction actively requesting help rather than help being provided in an unsolicited manner during their interaction with a simulation environment. Note that in contrast to other published results that consistently point towards problem solving as the most promising point on the Exploratory Learning Continuum, these results point in the opposite direction, towards a less strongly guided approach, although they do not explicitly evaluate these two approaches in comparison with problem solving.

In a series of classroom studies in which we have controlled both for content and for time on task, we have empirically evaluated our Negotiable Problem Solving Goals approach, which falls in between problem solving and the types of guided exploratory learning evaluated in the past (e.g., Leutner, 1993; Shute & Glaser, 1990). This approach is student centered in terms of allowing the student to make choices about the direction their exploration will take, and yet the tutor also plays an important role in shaping this path by looking for opportunities on the selected path to stimulate reflection, or occasionally discouraging students from going down a path they are not prepared to learn from yet. Thus, the student's learning path emerges as a product of the negotiation between the student and the tutor. In a first study (Rosé et al., 2005), we contrast three goal level conditions, namely Negotiable Problem Solving Goals (with students working with human tutors), tutorial learning (a reading control condition), and problem solving (implemented with the Cognitive Tutor Authoring Tools (CTAT) (Koedinger et al., 2004; Aleven & Rosé, 2005). In all three conditions, students interacted with a simulation environment. Overall there was a main effect for the Goal Level manipulation ($F(2,83) = 3.81$, $p < .05$, $MSE = 20.9$). Overall the order was PS < S < NPSG. Using a Bonferroni post-hoc analysis, we determined that the difference between NPSG and PS was significant ($p < .05$), whereas the difference between NPSG and S was marginal ($p=.11$). The difference between the S and PS was only a trend.

We then conducted two follow-up classroom evaluations in which we implemented the NPSG condition with tutorial dialogue agents (Gweon et al., 2005) rather than human tutors, with complete details reported in (Kumar et al., 2006). These studies demonstrated a significant learning benefit for the NPSG tutorial dialogue agents, and specifically an advantage for the implemented NPSG approach over the Script (S) condition from the original CycleTalk study, with an effect size of .25 standard deviations. In our proposed work, we will evaluate the generality of this basic approach with a new age group and domain, specifically middle school math learning. Our prior work related to the NPSG approach provides a technological framework for supporting student directed learning by actively eliciting reflection and offering guidance at key points on the student directed learning path. In our previous work we have demonstrated that computer agents can be successful at drawing out student explanations when they offer feedback that is directed at the details of what students say (Rosé & Torrey, 2005). The proposed work differs from our prior work with NPSG in that we will be using the dialogue agents to support collaboration between students, rather than as an educational intervention for individual learning with technology. We discuss our proposed research extending our current tutorial dialogue technology in Section 4.3.

The form of collaboration support we offer with our NPSG-based approach differs from more typical forms of collaborative learning support explored in the Computer Supported Collaborative Learning (CSCL) community in that it is adaptive, in other words responsive to the patterns of collaborative discourse as they unfold. We have already conducted a series of experimental studies in which we have accumulated some foundational knowledge on which to base the initial design for the adaptive collaboration support mechanism that we propose to optimize. In order to encourage productive patterns of collaborative discourse, researchers both in the CSCL tradition (Pfister and Mühlpfordt, 2002; Rummel, et al., 2003; Weinberger et al., in press) and the Educational Psychology tradition (O'Donnell and Dansereau, 1992; Palincsar and Brown, 1984; King, 1998) have separately developed approaches for scaffolding the interactions between students, to help them coordinate their communication, and to encourage deep thinking and elaborated explanations. These simple forms of support are implemented as "scripts" or static sets of prompts that students are provided with and expected to respond to. There is much evidence that argues for the effectiveness of these simple forms of support for boosting productive conversational behaviors (Pfister and Mühlpfordt, 2002; Rummel, et al., 2003; Weinberger et al., in press). However, we argue that in an on-line setting which students will interact with an environment over an extended period of time, a support mechanism that is not sensitive to what is happening in the collaboration will eventually be ignored.

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 adaptive support for on-line collaborative learning (Gweon et al., 2006). In an initial study we found that even without adaptive support, problem solving in pairs is significantly more effective for learning than problem solving alone, $F(1,18) = 6.0$, $p<.05$, $MSE = 5.64$, effect size = 1.1 standard deviations. Nevertheless, we found many individual differences between students in terms of their level of engagement with the problem solving and their relative ability level. Reflections from the first study lead us to ask what strategies for matching students with learning partners would produce the optimal conditions for learning. Previous work in this area provides mainly correlational evidence in favor of some combinations of ability levels and style over others. Building upon the results from the first study, we experimentally explored these group dynamics issues in greater depth in a second study in which we carefully manipulated engagement and ability level in confederate peer learners in order to precisely measure the causal impact of these variables on the behavior and learning of participant students working with them in pairs. We predicted an interaction between these variables since intuitively if a high ability and a low ability student are working together, for example, it makes sense for the high ability student to take the lead. The results from the second study offer limited evidence of the instructional value of exposure to errors. Beyond that, the results suggest that certain combinations of these two variables make students dangerous learning partners. We refer to these dangerous combinations as *the danger zone*.

The results from the second study motivate the design of adaptive prompts both to encourage deep explanation and teaching behavior and to manipulate student behavior in an attempt to keep it out of the danger zone identified in the second study. We evaluated this design in a third study in which we manipulate both the ability level of the confederate peer learner (as in Study 2) as well as contrasting a condition in which adaptive prompts are offered to one in which they are not offered. Prompts had a positive effect on student behavior in the intended direction, thus offering evidence in favor of our design for adaptive support. However we also observed some negative effects of prompts on student performance that also negatively interacted with student learning within the condition where students worked with a Low Ability confederate peer learner. In particular, there was a non-significant trend for distribution of labor prompts to reduce the number of correct problems solved within that condition. Although the effect was not significant, the added noise in terms of number of correct problems submitted obscured the difference in learning between the Prompt and No Prompt conditions. When we factor out this effect by including correct problems submitted as a covariate in an ANCOVA comparing pre to post test gains of students in the Prompt and No Prompt condition, we see a significant benefit for prompting on student learning. $F(1,39) = 4.12$, $p < .05$. These results show promise that adaptive collaboration support can have a positive impact on student learning in pairs while demonstrating that further experimentation is needed in order to optimize the design. In future iterations we plan to modify our distribution of labor prompts so that the ideal distribution of labor is dependent upon the relative ability levels of the two students. Further directions for experimentation are discussed in Section 3.

We have begun to experimentally explore issues related to the design of an effective environment for collaborative learning for middle school math. With middle school children in the math domain we have explored how content based feedback from a learning environment affects conversational interactions, perceived interdependence, and student learning (Gweon et al., submitted). We conducted an empirical investigation in which we experimentally contrasted collaboration in two feedback configurations, one which is identical to the state-of-the-art in intelligent tutoring technology, which we refer to as the Immediate Feedback condition, and one which is based on a long line of investigation of the use of worked out examples for instruction, which we refer to as the Delayed Feedback condition. We investigated the impact of this experimental manipulation on perceptions about the collaboration revealed by a questionnaire, evidence of learning from tests and quizzes, and a qualitative analysis of the collaborative problem solving process from coded chat logs collected during the collaborative problem solving sessions. We find a gender by condition interaction in which male students prefer and benefit more from the Immediate Feedback condition while female students prefer and benefit more from the Delayed Feedback condition. Nevertheless, what is in common between male and female students was that they perceived benefit as well as benefited to the extent that they offered help to their partner. Thus, similar to findings from (Gweon et al., 2006), we have strong evidence of the Learning by Teaching effect, and thus reason to believe that by eliciting more explanation and teaching behavior from students, even in this age group, we can increase the educational benefit that they receive and perceive.

In order to elicit the type of collaborative behavior that leads to more learning, we will use adaptive collaboration support in the style of our previous investigations at the college level (Gweon et al., 2006). Our previous success with automating collaborative learning process analysis (Donmez et al., 2005) offers promise that the adaptive support mechanism evaluated using a Wizard-of-Oz setup in (Gweon et al., 2006) can be implemented and deployed fully automatically. We have achieved near human reliability at applying a sophisticated multi-dimensional collaborative process analysis coding scheme describe in (Weinberger & Fischer, submitted; Fischer et al., 2002) to a corpus of newsgroup style collaborative learning interactions using our novel text classification technology. In the proposed work described below in Section 4.2 we discuss how we propose to build on and extend this technological foundation, which is a key enabling technology for our adaptive collaborative learning support approach.

3. Exploring the Design Space through Intervention Studies

Controlled investigations exploring the central behavioral research questions this proposal is concerned with will take place in the lab (at CMU). We will use a series of lab studies to refine our design for adaptive collaborative support before testing our most promising approaches in a classroom setting or on-line “in the wild”. Lab studies allow us to obtain larger sample sizes than are readily available in typical classroom or on-line learning settings within a reasonable amount of time and allow us to exert more experimental control.

Research questions that we address through our lab studies, which build on findings from our prior explorations of collaborative problem solving support, include the following:

- What supplementary materials should we provide students with to assist them in their help seeking and help providing behavior?
- When and how often should we intervene in the collaborative problem solving process?
- What types of prompts are most effective for supporting help seeking and help providing behavior?

3.1 Experimental Paradigm

All lab studies will be approved through CMU's Institutional Review Board (IRB) and will use the following experimental paradigm.

Participants. Participants will be middle school children recruited through local newspapers 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 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 statistical analyses. Furthermore, we will use multi-level modeling in order to control for pair effects in our between condition comparisons. Subject to formal parent consent and student assent, students will sign up to come to the lab in such a way that pairs of students will be scheduled to work together, but we will avoid signing up students working with other students that they know. We will attempt to accomplish this through phone interviews where we will collect information about where students go to school and what extra-curricular activities they are involved in.

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). Collaborative work sessions will take place in the VMT environment described in Section 4.1.

Experimental Procedure. We will strictly control for time in all 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 counter-balance the order of the tests in order to control for any potential differences in difficulty between tests. 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 Manipulatio*n. Based on our previous experience, with 20 pairs per condition, we expect each 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 (See Section 4.2), 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, in press; Strijbos, 2004; Lally & De Laat, 2002). Variables related to group process such as amount of deep explanation behavior, help seeking and help provision behavior, and measures of *transactivity* (as introduced in Section 2 and discussed further in Section 4.2 below) will be analyzed both as ends in themselves, i.e., examining the effect of our experimental manipulations on patterns of communication, but also as covariates 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.

3.2 Example study: Eliciting Precise Explanations

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. We can provide all of these things as resources to students in the spirit of the type of non-adaptive support for collaboration offered to students in state-of-the-art collaboration learning environments. These forms of support have already proven useful in prior studies of collaborative learning. 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., submitted), we observed that students sometimes realized they should offer help but were not able to.

Thus, our first lab study will be a 2X3 factorial design in which we will contrast Non-adaptive support versus No Non-adaptive support as one factor and Adaptive support (in the form of simple adaptive prompting as in (Gweon et al., 2006)) versus more elaborate Adaptive support by means of tutorial dialogue agents (See Section 4.3) versus No Adaptive support. 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 are in one of the Adaptive support conditions where they receive prompts to attempt to offer help. We hypothesize that the conditions in which students receive both the Adaptive support and the Non-adaptive support will be the conditions in which they learn the most and offer the most high quality help. We hypothesize that low ability students will benefit more from the more elaborate form of Adaptive support, whereas it will not be necessary for high ability students. Furthermore, we hypothesize that if we are successful in improving the quality of help offered, we will reduce the extent to which prompts that seek to keep the division of labor equal are detrimental to problem solving efficiency when there are large differences in ability level between students.

3.3 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 factor we would like to explore is whether our adaptive support should focus on 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. At a high level, the collaboration support we offer students is meant to increase the level of *transactivity* in the collaborative discourse. The manipulations we have discussed thus far all focus on overcoming problem solving difficulties. Another alternative would be to contrast support that focuses on overcoming problem solving difficulties versus support that simply focuses on clarity. For example, a prompt that focuses on clarity might simply request one student to rephrase what the other student has said. Support that encourages students to strive for greater clarity might have the effect of increasing help seeking and help providing behavior as well, but might lead to reflective discussions arising at different times in the problem solving process, such as during times when problem solving is progressing well rather than only when at least one student is struggling.

4. Building on and Extending the Technological Infrastructure

4.1 Collaborative Environment

The Math Forum and its Virtual Math Teams Project will collaborate with CMU personnel under this grant in the designing, developing, implementing, testing, refining and deploying of the computer-support tools that are part of this grant. In particular, the VMT-Chat environment will be available as a test-bed for collecting data about the performance of these tools. VMT staff will be involved in assessing the results of the use of these tools through close analysis of selected excerpts from this chat log data. Here we discuss the current VMT environment as well as what we propose to add to it. The free VMT service currently consists of an introductory web portal within the Math Forum site (http://mathforum.org/vmt) and an interactive environment called VMT-Chat. 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).

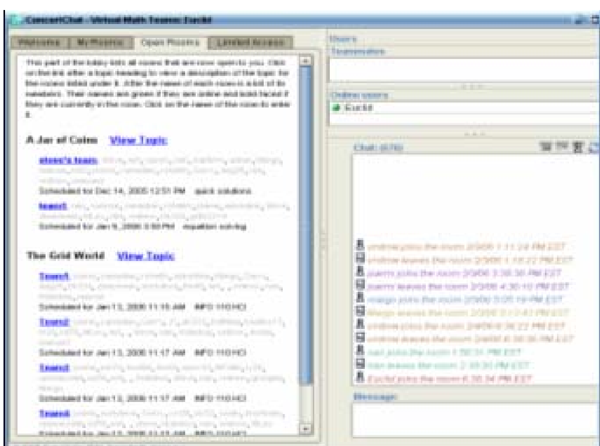

Figure 2. The VMT Lobby.

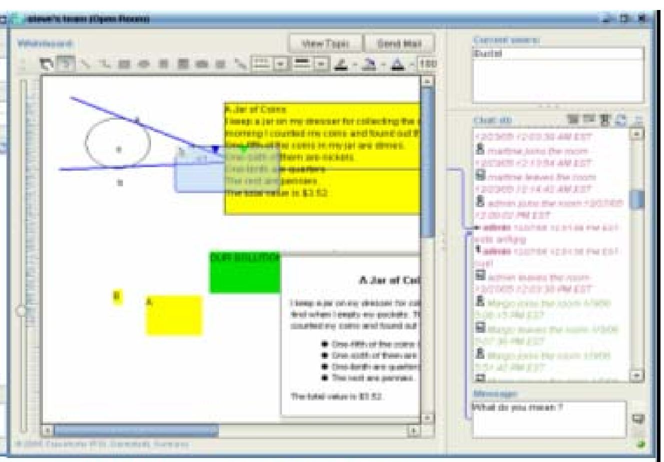

Figure 3. A VMT chat room.

Three types of rooms can be created in the lobby:

a. *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.

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

c. *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:

1. 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.
2. Later they could split up and join groups with students from other schools to explore more open-ended mathematical situations.
3. 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.

Through such sequences, people become more active members of a math discourse virtual community and help to grow that community. The Drexel Team will be responsible for integrating the technology described in Sections 4.2 and 4.3 with the VMT environment both for the purpose of enabling the lab studies described in Section 3 but also to integrate a fully-automatic form of adaptive collaboration support building both upon findings from lab studies as well as the results from technological development proposed under this grant in order to produce a supportive environment to test in classroom studies at Propel Charter School (See Section 6) and on-line in the actual VMT on-line service in the final year of the proposed 3 year project.

4.2 Detecting Opportunities for Intervention

	Dimension	Description	Number of Classes
Knowledge	Epistemic (EPI)	How learners work on the learning task, e.g., what content they are referring to or applying	35
Language	Micro (ATOL)	How an individual argument consists of a claim which can be supported by a ground with warrant and/or specified by a qualifier	4
	Macro (ALEI)	How argumentation sequences are examined with respect to how learners connect single arguments and create an argumentation pattern together	6
Relational Style	Social Modes (SOC)	To what degree or in what ways learners refer to the contributions of their learning partners	21
	Reaction (REA)	Coarser grained version of SOC	3
	Appropriateness (PRO)	How learners make use of prompts, i.e., whether learners could *use the prompts in the intended manner*, e.g. write an argument when they are asked to write a counterargument	4
	Quoted (QUO)	Distinguishes between new contributions and quoted contributions	2

Table 1 Coding scheme from (Fischer et al., 2002) used for characterizing collaborative learning process

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). Table 1 summarizes the multi-dimensional coding scheme that we achieved near human levels of reliability at automatically applying to newsgroup style collaborative learning interactions. We refer to this coding scheme, developed by Fischer and colleagues (2002), as the Tuebingen coding scheme. Note that each span of text is assigned multiple labels, specifically one code from each of the 7 dimensions displayed in Table 1, each of which offers a different perspective on the nature of the contribution, often drawing upon information of a different nature from the other dimensions. For example, the Micro and Macro dimensions each characterize different aspects of the linguistic structure of the contributions whereas the Social Modes, Reaction, and Appropriateness 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. Continued research on this problem is reported in (Donmez et al., submitted) where we develop a new ensemble-based approach to achieving increased reliability in multi-class classification that out-performs well established previous approaches such as boosting. On-line application of coding schemes such as the Tuebingen coding scheme can be used to trigger adaptive support for collaborative communication processes.

Our successful proof of concept was trained using a supervised machine learning approach, which means that we began with hand labeled data and trained models to apply the same codes to unseen data. This approach would be sufficient for triggering adaptive collaboration support within a limited domain, where the nature and content of the conversations are very similar to what is found in obtainable training data. In our proposed work, we seek to expand upon this foundation by developing a more general purpose automatic measure of collaborative process quality, which we have referred to in this proposal as *transactivity*. We draw from other prior work on collaborative learning process analysis (Meier et al., submitted; Weinberger & Fischer, in press; Strijbos, 2004; Lally & De Laat, 2002).

Our proposed approach builds upon our recent work on topic segmentation of dialogue (Arguello & Rosé, 2006a; Arguello & Rosé,2006b). In this work we have introduced a novel topic segmentation approach that combines evidence of topic shifts from lexical cohesion with linguistic evidence such as syntactically distinct features of segment initial contributions. Our evaluation demonstrates that this hybrid approach outperforms all previously published state-of-the-art algorithms for automatic topic segmentation (e.g., (Hearst, 1997; Foltz et al., 1998; Barzilay & Lee, 2004; Olney & Cai, 2005) when applied to loosely structured, spontaneous dialogue. Using a similar approach, we have also automated the detection of exchange boundaries (Sinclair & Coulthard, 1975). Basic building blocks of dialogue structure, including exchanges and topic segments, are important for assessing the level of transactivity in dialogues. For example, dialogues in which topic segments are long, and topics shift rarely and in a smooth manner, are more likely to be high in transactivity than dialogues where topics shift rapidly and frequently. Furthermore, a topic analysis is useful for separating out topic words from stylistic words, thus facilitating effective learning of stylistic constructions that might be indicative of important discourse moves related to argumentative knowledge construction. Recently much work in the text classification and information extraction community has focused on the development of what are called bootstrapping or semi-supervised methods for acquiring classification and extraction models (Rilof & Jones, 1999; Ko & Jungyun, 2004; Witten & Frank, 2005; Tur et al., 2005). These approaches dramatically reduce the amount of human effort required to learn or tailor models to new corpora. Building upon techniques we have applied to the problem of automatic topic segmentation, we plan to adopt a bootstrapping approach for acquiring patterns that match against stylistic elements that are indicative of the types of conversational actions that are predictive of the level of transactivity in collaborative dialogue.

Note that what we are proposing is different from the purely numeric coherence measures that are part of the Cohmetrix system (see http://cohmetrix.memphis.edu/cohmetrixpr/index.html) in that as part of our approach to evaluating transactivity we compute aspects of explicit structure, such as exchange boundaries, segment boundaries, and identification of stylistic elements that we can formally reason with, such as to determine which speaker has conversational control. We seek to emulate the approach taking in the Tuebingen coding scheme where transactivity is evaluated in terms of distinct consensus building styles rather than by assigning a numeric value along a unidimensional scale.

4.3 Delivering Interventions

In our current work we are building an infrastructure to support quick authoring of dialogue agents (Gweon et al., 2005). Current objectives include building 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. 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.

In the proposed work, we will build on this foundation by using this technology to author interventions such as the Elaborate Adaptive support described in Section 3.2. While our previous work developing dialogue agents has focused on tutorial dialogue for individual learning, here we expand our scope to cover tutorial dialogue for collaborative learning. 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 the interactive interventions described in Section 3.2. The implemented models will differ from directed lines of reasoning in that the language understanding models that are used to process the elicited explanations must be adapted to the context in which the support was triggered.

5. Evaluating Adaptive Collaboration Support in the Classroom and in the Wild

The computer-based tools developed under the proposed grant will be tested in real-world use situations in the 3rd year of the proposed 3 year project both at Propel Charter School and on-line at the Math Forum. Specifically, these tools will be integrated with the VMT-Chat online collaboration environment. Regular uses of the VMT-Chat environment by middle-school students will take place while the environment is instrumented with these tools.

The tools will be used in three 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 Section 4.2 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. Depending upon success, in a limited number of trial cases, 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. This is also the style in which the environment will be tested at

Propel Charter School during lab sessions associated with specific math lessons related to content used in the lab studies at CMU.

While the classroom study at Propel Charter School will be a controlled study, explorations in the VMT environment will be more naturalistic. Analysis of the naturalistic trial cases will mainly take the form of case studies. 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 qualitative observations to complement the more quantitative findings from our controlled investigations. Their value comes from the highly externally valid insights we will gain from them in terms of moving towards our goal of providing a high quality on-line learning environment at a very low cost.

6. Partnerships

We are partnering with Propel Charter School in Homestead, Pennsylvania where we will run a Math Camp this summer in order to collect data on math explanations from our target user population in connection with the specific materials we plan to use in our lab studies and where we will run a classroom study in the final year of our proposed three year project. See letter of support from Propel Charter School's math coach, Ariane Watson.

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, primarily algebra and geometry. This site is well established. 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. The site offers a wealth of problems and puzzles; online mentoring; research; team problem solving; collaborations; and professional development. 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.

The Math Forum offers a number of online services, including but not limited to the following:

a. *The Problem of the Week*. This popular service posts a different problem every other week during the school year in a number of categories, such as math fundamentals, pre-algebra, algebra, geometry. Challenging non-standard math problems can be answered online or offline, with the opportunity for feedback from mentors on problem-solving and communication skills.

b. *Ask Dr. Math*. Students and others receive mathematics advice from professionals and expert volunteers.

c. *Math Tools*. Visitors to the site explore the world of interactive tools for understanding math concepts and communicate with teachers using them in their classrooms, discussing and rating the tools.

7. Results from Prior NSF Funding

Rosé has supervised 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 on adaptive collaboration support that this proposal is built upon. This project began by exploring the idea of the instructional benefit of errors, which has its roots in Piaget's notion of *perturbation* or *cognitive conflict* (Piaget, 1985). This cognitive conflict plays an important role in stimulating cognitive restructuring by making children aware of a deficiency in their current understanding for explaining the world around them. In a recent study (Gweon et al., 2005) we reexamined the effects of group composition on the functioning of collaborative learning dyads in the light of recent work on learning from incorrectly worked examples (Grosse and Renkl, submitted). In a follow-up study we explored the use of prompts to encourage more teaching oriented behavior from the student participants in light of results indicating that students benefit more from working with less capable peers when they engage in deep explanation activities (Gweon et al., 2006). This study demonstrated that adaptive support for collaboration increases teaching behavior and has a significant positive effect on student learning. Other publications from this work include (Gweon et al., 2005b) and (Rosé & Torrey, 2005). The Calculategy project heavily involved Master's students in human-computer interaction and language technologies, two of whom have recently been accepted into doctoral programs in prestigious schools and plan to pursue research in the area of educational technology development.

Weimer 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. See http://www.mathforum.org/vmt/researchers/orientation.html for more information as well as an extensive set of publications originating from this work.

8. Timeline and Management Plan

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 studies and the classroom study in the final year of the project. Weimer and Stahl will oversee all work conducted at Drexel University, which includes integrating adaptive collaboration support technology with the VMT environment and conducting naturalistic observations on-line in the VMT environment. The CMU and Drexel team will conduct phone conferences twice a month to coordinate their efforts. They will collaborate closely on the development of the materials for the studies as well as the analysis of data from all lab, classroom, and on-line studies and observations.

In year 1 of the proposed project, work on automatic process analysis and interactive support delivery will begin at CMU, and the VMT environment will be outfitted to begin lab studies at CMU. The first lab study, described in Section 3.2, will be conducted in the second half of the first year of the project. In the second year, two lab studies will be conducted at CMU. The bulk of the work on basic technology research at CMU will be completed and integrated with the VMT environment by the end of the second year of the project. In the final year of the project, naturalistic on-line observations in the VMT environment will begin at Drexel University. At CMU, a final lab study will be completed in addition to small improvements to the basic adaptive collaboration support technology, which will be integrated with the VMT environment by the mid-point of the final year of the project. In the second half of the final year of the project, more extensive naturalistic evaluations of the new VMT environment will be conducted at Drexel University. A final lab study will be conducted at CMU in parallel with a classroom study taking place at Propel Charter School. 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. Synergistic with the proposed work, graduate courses in Computer Supported Collaborative Learning are offered by Rosé at CMU and Stahl at Drexel University. Specifically in the CMU course, semester long group projects related to the design and implementation of prototype adaptive collaboration support mechanisms are a key component of the course.

Aleven, V, & Rosé, C. P. (2005). Authoring plug-in tutor agents by demonstration: Rapid tutor development, *Proceedings of AI in Education '05*.

Anderson, J.R., Boyle, C.F., Corbett, A., Lewis, M.W. Cognitive modeling and intelligent tutoring. *Artificial Intelligence*, 42, (1990), 7-49.

Arguello, J. & Rosé, C. P. (2006). Museli: A Multi-source Evidence Integration Approach to Topic Segmentation of Spontaneous Dialogue, Proceedings of *the North American Chapter of the Association for Computational Linguistics*

Arguello, J. & Rosé, C. P. (2006b). InfoMagnets: Making Sense of Corpus Data. Companion Proceedings of the North American Chapter of the Association for Computational Linguistics.

Regina Barzilay and Lillian Lee (2004). Catching the drift: Probabilistic Content Models, with Applications to Generation and Summarization. In *Proceedings of HLT-NAACL*, pp 113 – 120.

Benzmueller, C., Fiedler, A., Gabsdil, M., Horacek, H., Kruijff-Korbayova, I., Pinkal, M., Siekmann, J., Tsovaltzi, D., Quoc Vo, B., & Wolska, M. (2003). A Wizard-of-Oz Experiment for Tutorial Dialogues in Mathematics. Paper presented at the aied2003: Workshop on Advanced Technologies for Mathematics Education, 2003, Sydney, Australia.

Bloom, B. S. (1984). The 2 Sigma Problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13,* 4-16.

Brandon, P. R., & Newton, B. J. (1985). The superiority of girls over boys in mathematics achievement in Hawaii. *A presentation at the annual conference of the American Education Research Association*, Chicago, April. (ED 260Ê906)

Cakir, M., Xhafa, F., Zhou, N., & Stahl, G. (2005). *Thread-based analysis of patterns of collaborative interaction in chat.* Paper presented at the international conference on AI in Education (AI-Ed 2005), Amsterdam, Netherlands. Retrieved from http://www.cis.drexel.edu/faculty/gerry/pub/aied2005.pdf.

Callahan, L. G., & Clements, D. H. (1984). Sex differences in rote-counting ability on entry to first grade: Some observations. *Journal for Research in Mathematics Education*, 15, 378-382.

Chapin, S., O'Connor, C., & Anderson, N. (2003). *Classroom Discussions: Using Math Talk to Help Students Learn*, Math Solutions Publications: Sausalito, CA.

Charney, D.H., & Reder, L.M. (1986). Designing tutorials for computer users: Effects of the form and spacing of practice on skill learning. *Human Computer Interaction*, 2, 297-317.

Chi, M. T. H. (2000). Self-Explaining Expository Texts: The Dual Processes of Generating Inferences and Repairing Mental Models. In R. Glaser (Ed.), *Advances in Instructional Psychology,* (pp. 161-237). Mahwah, NJ: Erlbaum.

Chi, M. T. H., de Leeuw, N., Chiu, M. H., LaVancher, C., (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*(3), 439-477.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., and Glaser, R., (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, 13:2, 145-182.

Cohen, P. A., Kulik, J. A., & Kulik, C. C. (1982). Educational Outcomes of Tutoring: A meta-analysis of Findings. *Americal Educational Research Journal, 19,* 237-248.

De Lisi, R. & Goldbeck, S. L. (1999). Implications of Piagetian Theory for Peer Learning, in O'Donnell & King (Eds.) *Cognitive Perspectives on Peer Learning*, Lawrence Erlbaum Associates: new Jersey.

Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher, 32* (1), 5-8.

Donmez, P., Carbonnel, J., Rosé, C. P. (submitted). Title: TurboCharging: A New Cascaded Ensemble Learning Method. Submitted to the European Conference on Machine Learning.

Donmez, P., Rose, C. P., Stegmann, K., Weinberger, A., and Fischer, F. (2005). Supporting CSCL with Automatic Corpus Analysis Technology, to appear in *the Proceedings of Computer Supported Collaborative Learning*.

Dossey, J. A., Mulis, I. V. S., Lindquist, M. M., & Chambers, D. L. (1988). *The mathematics report card: Are we measuring up: Trends and achievement based on the 1986 National Assessment.* Princeton: Educational Testing Service.

P.W. Foltz, W. Kintsch, and Thomas Landauer (1998). The measurement of textual cohesion with latent semantic analysis. *Discourse Processes*, 25, 285-307

Grosse, C. & Renkl, A. (submitted). Finding and Fixing Errors in Worked Examples: Can this foster learning outcomes? Draft submitted for review for journal publication.

Gweon, G., Rosé, C. P., Cui, J., Albright, E. (submitted). Towards the design of a collaborative problem solving environment that promotes positive interdependence. Submitted to the Computer Supported Cooperative Work Conference.

Gweon, G., Rosé, C. P., Zaiss, Z., & Carey, R. (2006). Providing Support for Adaptive Scripting in an On-Line Collaborative Learning Environment, Proceedings of *CHI 06: ACM conference on human factors in computer systems.* New York: ACM Press.

Gweon, G., Rosé, C. P., Wittwer, J., Nueckles, M. (2005). An Adaptive Interface that Facilitates Reliable Content Analysis of Corpus Data, *Proceedings of Interact '05*

Gweon, G., Arguello, J., Pai, C., Carey, R., Zaiss, Z., Rosé, C. P. (2005). Towards a Prototyping Tool for Behavior Oriented Authoring of Conversational Interfaces, *Proceedings of the ACL Workshop on Educational Applications of NLP.*

Gweon, G., Rosé, C. P., Carey, R., Zaiss, Z. (2005). Towards Data Driven Design of a Peer Collaborative Agent, *Proceedings of AI in Education '05*

Marti Hearst. 1997. TextTiling: Segmenting Text into Multi-Paragragh Subtopic Passages. *Computational Linguistics*, 23(1), 33 – 64.

Hutchins, E. (1996). *Cognition in the wild*. Cambridge, MA: MIT Press.

King, A. (1999). Discourse Patterns for Mediating Peer Learning, in O'Donnell & King (Eds.) *Cognitive Perspectives on Peer Learning*, Lawrence Erlbaum: New Jersey.

King, A. (1998). Transactive peer tutoring: Distributing cognition and metacognition. Computer-Supported Cooperation Scripts 52 Educational Psychology Rev. 10: 57-74.

Klahr & Nigam (2004). The equivalence of learning paths in early science instruction: effects of direct instruction and discovery learning, *Psychological Science*, 2004.

Ko, Y. & Jungyun, S. (2004). Learning with Unlabeled Data for Text Categorization Using Bootstrapping and Feature Projection Techniques, *Proceedings of the Association for Computational Linguistics.*

Koedinger, K. R., Aleven, V., Heffernan, N., McLaren, B., & Hockenberry, M. (2004). Opening the Door to Non-Programmers: Authoring Intelligent Tutor Behavior by Demonstration. In J. C. Lester, R. M. Vicario, & F. Paraguaçu (Eds.), *Proceedings of Seventh International Conference on Intelligent Tutoring Systems, ITS 2004* (pp. 162-174). Berlin: Springer Verlag

Koschmann, T., Stahl, G., & Zemel, A. (2006). The video analyst's manifesto (or the implications of Garfinkel's policies for the development of a program of video analytic research within the learning sciences). In R. Goldman, R. Pea, B. Barron & S. Derry (Eds.), *Video research in the learning sciences*. Retrieved from http://www.cis.drexel.edu/faculty/gerry/publications/journals/manifesto.pdf.

Kumar, R., Rosé, C. P., Aleven, V., Iglesias, A., Robinson, A. (2006). Evaluating the Effectiveness of Tutorial Dialogue Instruction in an Exploratory Learning Context, *Proceedings of the Intelligent Tutoring Systems Conference.*

Lally, Vic and De Laat, Maarten F. (2002) Cracking the code:Learning to collaborate and collaborating to learn in a networked environment. In, CSCL, University of Colorado, Boulder, USA, 7-11 January, 2002.

Thomas Landauer, & Susan Dumais (1997). A Solution to Plato's Problem: The Latent Semantic Analysis of Acquisition, Induction, and Representation of Knowledge. *Psychological Review*, 104, 221-240.

Leutner, D. (1993). Guided discovery learning with computer-based simulation games: effects of adaptive and non-adaptive instructional support. *Learning and Instruction*, 3, 113-132.

Maxwell, J. (2004). Causal explanation, qualitative research, and scientific inquiry in education. *Educational Researcher*, *33* (2), 3-11. Retrieved from http://www.aera.net/pubs/er/pdf/vol33_02/2026-02_pp03-11.pdf.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The Case for Guided Methods of Instruction, *American Psychologist* 59(1), pp 14-19.

Meier, A., Spada, H., Rummel, N. (submitted). A Rating Scheme for Assessing the Quality of Computer-Supported Collaboration Processes, Submitted to *the International Journal of Computer Supported Collaborative Learning*

Meloth, M. S. & Deering, P. D. (1999). The Role of the Teacher in Promoting Cognitive Processing During Collaborative Learning, in O'Donnell & King (Eds.) *Cognitive Perspectives on Peer Learning*, Lawrence Erlbaum Associates: new Jersey.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Gregory, K. D., Garden, R. A., O'Connor, K. M., Chrostowski, S. J., & Smith, T. A. (2000). *TIMSS 1999 International Mathematics Report.* Boston, MA: The International Study Center, Boston College, Lynch School of Education.

O'Donnell, A. M. (1999). Structuring Dyadic Interaction Through Scripted Cooperation , in O'Donnell & King (Eds.) *Cognitive Perspectives on Peer Learning*, Lawrence Erlbaum Associates: new Jersey.

O'Donnell, A. M., and Dansereau, D. F. (1992). Scripted Cooperation in student dyads: A method for analyzing and enhancing academic learning and performance. In Hertz-Lazarowitz, R., and Miller, N. (Eds.), Interaction in cooperative groups: The theoretical anatomy of group learning, Cambridge University Press, New York, pp. 120-141.

Andrew Olney and Zhiqiang Cai (2005). An Orthonormal Basis for Topic Segmentation of Tutorial Dialogue. *Proceedings of HLT-EMNLP'05.*

Palincsar, A. S., and Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. Cognition and Instruction 1: 117-175.

Piaget, J. The equilibrium of cognitive structures: the central problem of intellectual development, Chicago University Press, 1985.

Pfister, H.-R., and Mühlpfordt, M. (2002). Supporting discourse in a synchronous learning environment: The learning protocol approach. In Stahl, G. (Ed.), Proceedings of the Conference on Computer Supported Collaborative Learning (CSCL) 2002, Erlbaum, Hillsdale, pp. 581-589.

Renninger, K. A., & Shumar, W. (2002). *Building virtual communities*. Cambridge, UK: Cambridge University Press.

Riloff, E. & Jones, R. (1999). Learning Dictionaries for Information Extraction by Multi-Level Bootstrapping, *Proceedings of the Sixteenth National Conference on Artificial Intelligence*.

Rosé, C. P., Kumar, R., Aleven, V., Robinson, A., Wu, C. (in press). CycleTalk: Data Driven Design of Support for Simulation Based Learning, *International Journal of Artificial Intelligence in Education Special Issue on The Best of ITS '04*

Rosé, C. P., Aleven, V., Carey, R., Robinson, A., Wu, C. (2005). A First Evaluation of the Instructional Value of Negotiatble Problem Solving Goals on the Exploratory Learning Continuum, *Proceedings of AI in Eduction '05*

Rosé, C., Donmez, P., Gweon, G., Knight, A., Junker, B., Cohen, W., Koedinger, K., & Heffernan, N (2005). Automatic and Semi-Automatic Skill Coding with a View Towards Supporting On-Line Assessment, *Proceedings of AI in Education '05.*

Rosé, C. P., & Torrey, C. (2005). Interactivity versus Expectation: Eliciting Learning Oriented Behavior with Tutorial Dialogue Systems, *Proceedings of Interact '05*

Rosé, C. P., Pai, C., Arguello, J. (2005). Enabling Non-linguists to Author Conversational Interfaces Easily, *Proceedings of FLAIRS 05.*

Rosé C. P., & VanLehn, K. (2005). An Evaluation of a Hybrid Language Understanding Approach for Robust Selection of Tutoring Goals, *International Journal of AI in Education 15(4).*

Rosé, C. P. & Hall, B. S. (2004). A Little Goes a Long Way: Quick Authoring of Semantic Knowledge Sources for Interpretation. *Proceedings of the Second International Workshop on Scalable Natural Language Understanding*.

Rosé, C. P., Gaydos, A., Hall, B., Roque, A., VanLehn, K. (2003), Overcoming the Knowledge Engineering Bottleneck for Understanding Student Language Input, *Proceedings of AI in Education*, 2003

Rosé, C. P., Roque, A., Bhembe, D., VanLehn, K. (2002). An Efficient Incremental Architecture for Robust Interpretation, *Proceedings of Human Languages Technologies Conference*, San Diego, California

Rosé, C. P., Jordan, P., Ringenberg, M., Siler, S., VanLehn, K., Weinstein, A. (2001a). Interactive Conceptual Tutoring in Atlas-Andes, *Proceedings of AI in Education* 2001

Rosé, C. P. (2000). A Framework for Robust Semantic Interpretation, *Proceedings of 1st Meeting of the North American Chapter of the Association for Computational Linguistics*

Rosé, C. P., Di Eugenio, B., Levin, L. S., Van Ess-Dykema, C. (1995). Discourse Processing of Dialogues with Multiple Threads , *Proceedings of the Association for Computational Linguistics*

Rummel, N., Spada, H., Caspar, F., Ophoff, J. G., Schornstein, K. Instructional support for computer-mediated collaboration – results from process analyses. *In Proc. CSCL* (2003), 199-208.

Sarmiento, J., Trausan-Matu, S., & Stahl, G. (2005). *Dialogical perspectives on narratives in collaborative mathematics problem-solving.* Paper presented at the International Symposium on Organizational Learning and Knowledge Work Management (OL-KWM 2005), Bucharest, Romania. Proceedings pp. 88-99. Retrieved from http://www.cis.drexel.edu/faculty/gerry/pub/romania.pdf.

Shute, Valerie J. and Glaser, Robert, (1990). "A Large Scale Evaluation Of An Intelligent Discovery World: Smithtown", *Interactive Learning Environments* 1, 51-77

Stahl, G. (2006a). Analyzing and designing the group cognitive experience. *International Journal of Cooperative Information Systems (IJCIS)*. Retrieved from http://www.cis.drexel.edu/faculty/gerry/pub/ijcis.pdf.

Stahl, G. (2006b). Group cognition in an online chat community: Analyzing collaborative use of a cognitive tool. *Journal of Educational Computing Research (JECR) special issue on Cognitive tools for collaborative communities*. Retrieved from http://www.cis.drexel.edu/faculty/gerry/pub/jecr.pdf.

Stahl, G. (2006c). *Group cognition: Computer support for building collaborative knowledge*. Cambridge, MA: MIT Press. Retrieved from http://www.cis.drexel.edu/faculty/gerry/mit/.

Stahl, G. (2006d). Scripting group cognition: The problem of guiding situated collaboration. In F. Fischer, H. Mandl, J. Haake & I. Kollar (Eds.), *Scripting computer-supported collaborative learning: Cognitive, computational and educational perspectives*. Dodrecht, Netherlands: Kluwer-Springer Verlag. Retrieved from http://www.cis.drexel.edu/faculty/gerry/pub/scripting.pdf.

Stahl, G. (2006e). Sustaining group cognition in a math chat environment. *Research and Practice in Technology Enhanced Learning (RPTEL), 1* (2). Retrieved from http://www.cis.drexel.edu/faculty/gerry/pub/rptel.pdf.

Strijbos, J. W., & Stahl, G. (2005). *Chat-based problem solving in small groups: Developing a multi-dimensional coding scheme.* Paper presented at the Eleventh Biannual Conference of the European Association for Research in Learning and Instruction (EARLI 2005), Nicosia, Cyprus. Retrieved from http://www.cis.drexel.edu/faculty/gerry/pub/earli2005jw.pdf.

Strijbos, J. W. (2004). The effect of roles on computer supported collaborative learning, Open Universiteit Nederland, Heerlen, The Netherlands

Teasley, S. (1995). The Role of Talk in Children's Peer Collaborations, *Developmental Psychology* 31(2), pp 207-220.

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.

Tuovinen, J. E., & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology*, 91(2), 334-341.

Tur, G., Hakkani-Tuer, D., & Schappire, R. (2005). Combining active and semi-supervised learning for spoken language understanding, *Speech Communication* 45, pp 171-186.

Wang, H., Li, T., Huang, C., Chang, C., Rosé, C. P. (2006). VIBRANT: A Brainstorming Agent for Computer Supported Creative Problem Solving, *Proceedings of the Intelligent Tutoring Systems Conference (poster).*

Webb, N., Nemer, K., & Zuniga, S. (2002). Short Circuits or Superconductors? Effects of Group Composition on High-Achieving Students' Science Assessment Performance, *American Educational Research Journal*, 39, 4, pp 943-989.

Webb, N. & Mastergeorge, A. (2003). The Development of Students' Helping Behavior and Learning in Peer-Directed Small Groups, Cognition and Instruction, 21(4), pp 362-428.

Webb, N. M. (1991). Task related verbal interaction and mathematics learningin small groups. *Journal for Research in Mathematics Education, 22*(5), 366-389.

Weinberger, A., Reiserer, M., Ertl, B., Fischer, F., and Mandl, H. (in press). Facilitating collaborative knowledge construction in computer-mediated learning with structuring tools. In Bromme, R., Hesse, F., and Spada, H. (Eds.), Barriers and Biases in networkbased knowledge communication in groups, Kluwer, Dordrecht.

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

Wessner, M., Shumar, W., Stahl, G., Sarmiento, J., Muehlpfordt, M., & Weimar, S. (2006). *Designing an online service for a Math community.* Paper presented at the International Conference of the Learning Sciences (ICLS 2006), Bloomington, IN. Retrieved from http://www.cis.drexel.edu/faculty/gerry/pub/icls2006design.pdf.

Witten, I. H. & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*, second edition, Elsevier: San Francisco, ISBN 0-12-088407-0

Zemel, A., Xhafa, F., & Cakir, M. (2005). *What's in the mix? Combining coding and conversation analysis to investigate chat-based problem-solving.* Paper presented at the 11th Biennial Conference of the European Association for Research on Learning and Instruction (EARLI 2005), Nicosia, Cyprus.