Enhancing Scientific Reasoning and Explanation Skills with Conversational Agents

Gregory Dyke, David Adamson, Iris Howley, and Carolyn Penstein Rosé

Abstract—This paper investigates the use of conversational agents to scaffold on-line collaborative learning discussions through an approach called Academically Productive Talk (APT). In contrast to past work on dynamic support for collaborative learning, which has involved using agents to elevate the conceptual depth of collaborative discussion by leading students in groups through directed lines of reasoning, this APT based approach lets students follow their own lines of reasoning and promotes productive practices such as deep explaination and refinement of ideas. The study provides evidence that one form of the support, namely the Revoicing support, lead to significantly more learning within an online collaborative learning activity and resulted in higher quality explanations for predictions within that session.

Index Terms—Education, Lingusitics, Psychology

1 Introduction

THE frequent occurrence of over-full classroom settings in schools has lead the classroom discourse community to question how discussions in classrooms can be academically productive, particularly if we wish to use such situations to develop scientific reasoning skills. A large body of work has shown that certain forms of classroom interaction, termed Accountable Talk, or *Academically Productive Talk* (APT), are beneficial for learning with understanding in subjects such as math and science [15,17]. In this paper we explore how we can achieve some of the benefits of this form of learning support within small online learning groups engaged in learning scientific content supported by technology.

In prior work using intelligent conversational agents to support collaborative learning, the agents have provided social support, affording the agents a more credible social standing in the group and helping to diffuse tension and create a productive learning environment. Furthermore, they have provided conceptual support, designed to elicit more depth by leading students through directed lines of reasoning, referred to as *knowledge construction dialogues* (KCDs). While KCDs have been shown to lead to increased learning gains in science [18], math [10] and Engineering [11], particularly in situations where the conversational agents also provide social support [1,2,12], the necessity of designing them statically, with a pre-defined line of reasoning in mind both makes them hard to adapt to new subject material and does not fully exploit the benefits of collaborative learners following their own spontaneous lines of reasoning.

• *F.A. Author is with the National Institute of Standards and Technology, Boulder, CO 80305. E-mail: author@ boulder.nist.gov.*
• *S.B. Author Jr. is with the Department of Physics, Colorado State University, Fort Collins, CO 80523. E-mail: author@colostate.edu.*
• *T.C. Author is with the Electrical Engineering Department, University of Colorado, Boulder, CO 80309. On leave from the National Research Institute for Metals, Tsukuba, Japan E-mail: author@nrim.go.jp.*

Manuscript received (insert date of submission if desired). Please note that all acknowledgments should be placed at the end of the paper, before the bibliography.

We have therefore drawn on and integrated extensive work related to support of classroom discourse to investigate the use by conversational agents of facilitation moves that promote Academically Productive Talk. The aim of APT facilitation moves is to increase the amount of displayed reasoning and *transactivity* [3], by dynamically reacting to student discussions, encouraging them to build on each other's reasoning. Furthermore, as APT refers both to learners social positioning with respect to each other and their conceptual positioning with respect to knowledge, this provides us with a theoretical framework to better integrate the social and conceptual support aspects of conversational agents.

In this paper, we present a first successful study involving an agent performing APT moves in the context of a 9^{th} grade biology classroom in an urban US school district. In the remainder of the paper, we first discuss the theoretical foundation for our work from the classroom discourse and computer supported collaborative learning communities. We then describe a new architecture for enabling the development of a new form of APT based dynamic collaborative learning support. Finally, we describe a classroom study involving students from 7 9^{th} grade biology classrooms that provides significant evidence in favor of one form of APT based support.

2 Theoretical Framework

The theoretical foundation for the work reported in this paper come from three areas. Specifically, we first draw from the literature on Academically Productive Talk. Next we draw from the literature on scripted collaboration from the Computer Supported Collaborative Learning community. Finally, we draw from the recent literature on Dynamic Support for Collaborative Learning.

2.1 Academically Productive Talk

The notion of Academically Productive Talk stems from frameworks that emphasize the importance of social interaction in the development of mental processes, and has developed in parallel to similar ideas from the computer-supported collaborative learning community. Michaels, O'Connor and Resnick (2007) describe some of the core dialogic practices of Accountable Talk along three broad dimensions [15]:

1. Students should be accountable to the learning community, listening to the contributions of others and building on them to form their own.
2. Students should be accountable to accepted standards of reasoning, emphasizing logical connections and drawing reasonable conclusions.
3. Students should be accountable to knowledge, making arguments which are based explicitly on facts, written texts or other public information.

Such practices are often unfamiliar in the classroom. Not only must they be introduced to students but it is necessary to provide teachers with the means to scaffold these interaction forms. Drawing on over 15 years of observation and study, Michaels, O'Connor and Resnick propose a number of core "moves" that teachers can draw upon in order to encourage the development of academically productive classroom discussion, among which are:

1. Revoicing: "So let me see if I've got your thinking right. You're saying XXX?" (with time for students to accept or reject the teacher's formulation);
2. Asking students to restate someone else's reasoning: "Can you repeat what he just said in your own words?";
3. Asking students to apply their own reasoning to someone else's reasoning: "Do you agree or disagree and why?";
4. Prompting students for participation: "Would someone like to add on?";
5. Asking students to explicate their reasoning: "Why do you think that?" or "How did you arrive at that answer?" or "Say more about that".

2.2 Script Based Support for Collaborative Learning

The Computer Supported Collaborative Learning community shares many of the same values related to desired conversational practices in student group discussions. What is different is that a teacher is normally not present to support those practices. Thus, it is necessary to design environments with affordances that play the same role, to whatever extent is possible. The most popular approach to providing such affordances in the past decade has been that of script based collaboration [4].

A script may describe any of a wide range of features of collaborative activities, including its tasks, timing, the distribution of roles, and the methods and patterns of interaction between the participants. A number of models have been proposed to aid in the design, description, and analysis of these rich models of collaboration [8,9,21,22]. Scripts can be classified as either macro-scripts or micro-scripts [6]. Macro-scripts are pedagogical models that describe coarsegrained features of a collaborative setting, which sequence and structure each phase of a group's activities to foster collaboration. Micro-scripts, in contrast, are models of dialogue and argumentation that are embedded in the environment, and are intended to be adopted and progressively internalized by the participants.

Examples of macro-scripts include the classic Jigsaw activity, as well as more tailored approaches like ArgueGraph and ConceptGrid [8]. Micro-scripting can be implemented by offering prompts or hints to the user to guide their contributions [20], which may depend on the current phase of the macro-script.

2.3 Dynamic Script Based Support With Conversational Agents

Early approaches to scripting have been static, offering the same script or supports for every group in every context. Such non-adaptive approaches can lead to over scripting [4], or to the interference between different types of scripts [22]. A more dynamic approach that triggers micro-scripted supports or the appropriate phases of macro-scripts in response to the automatic analysis of participant activity [19] would be preferable. This analysis could occur at a macro-discourse level, following the state of the activity as a whole, or it could be based on the classification of individual user contributions. Such dynamic awareness might allow minimal scripting to be used to greater effect, with greater hopes of the users internalizing the support's intended interaction patterns. Further, the benefits of fading the support over time [20] could be more fully realized, as the timing and degree of such fading could be tuned to the group's level of internalization. The collaborative tutoring agents described by Kumar and Rosé [13] were among the first to implement dynamic scripting in a CSCL environment.

Participants in a collaborative session aren't just completing the assigned task – they're involved in numerous simultaneous processes including social bonding, idea formation, argumentation, time management, and off-task activity. To allow for rich, interactive support of the whole interaction, a tutor must be able to express several differently-scoped behaviors concurrently - it can be considered to be working through several overlapping macro- and micro-scripts at once.

However, the tutor has to avoid looking silly or incompetent while doing so. A tutor managing several scripts at once can "step on its own toes". When multiple responses from the tutor interfere with, contradict, or interrupt each other, the tutor's illusion of competence is shattered, and it is subject to derision and abuse by all but the most polite of students. Although several approaches have been described to address some of these concerns [13], it has remained an actively-pursued grail [14]. The Bazaar architecture addresses this challenge with a modular framework for designing multi-party collaborative agents. Bazaar adapts the Basilica architecture to accommodate conflicting agent behavior, and offers an extensible mechanism for prioritizing and selecting proposed agent actions.

Bazaar is implemented as a core set of Java classes, plus a library of reusable behavioral components. Both the agent's overall composition and the configuration of each component are specified in plain-text properties files, offering the sort of low-overhead exibility for authoring, content, and deployment that recent work has championed, such as by Dillenbourg and Tchounikine [5]. Bazaar agents are able send and receive events from a varied set of collaborative environments, including ConcertChat [16], a text chatroom with a shared whiteboard, as well as novel environments like the virtual world of SecondLife. In the following section, we outline the way Bazaar addresses some shortcomings in the earlier Basilica architecture.

3.1 Prioritizing Proposed Actions

The various components of a Bazaar agent can propose user-facing actions in response to system events or student input – these are queued and managed by the agent's output coordinator, which periodically selects the highest-priority proposal, forwarding it to the environment to be enacted.

As a solution to the multi-policy coordination problem described in Section 2.3, we allow an extensible set of soft control strategies, based on the approach described by Lison [14]. A previously accepted agent action can leave a lingering presence with Bazaar's output coordinator, which can re-prioritize (or entirely suppress) incoming proposals until its influence expires. Each proposal is created with a timeout, after which it is no longer relevant - if a queued proposal has not been accepted when it expires, it is removed from the queue.

3.2 Automatic Revoicing

One of the forms of support evaluated in this paper is a Bazaar agent that performs a form of APT referred to as Revoicing. The agent compared student input (during the discussion of each cell-model experiment) against a list of correct statements drawn from the data collected in a pilot run of this study from the previous year. If an entry in this list could be interpreted as a paraphrase of the student's input (using methods similar to those described in Fernando and Stevenson's earlier work [7], it was offered by the agent as a "revoicing" to the students. Some examples are given in Table 1 below. The same statement was never offered more than once in the same session as a revoicing. When student statements were not close enough to match the revoicing list but contained the first mention of important lesson concepts (like "indicator" or "molecule size"), the agent would nudge the student or a peer to expand or restate their contribution.

The revoicing was offered by the agent in tandem with other forms of support not associated with this specific manipulation in the experiment using Bazaar's coordination ability described in Section 3.1. For example, the enactment of a revoicing response softly blocked any of the lower-priority social prompts that were triggered, until several seconds after the revoicing move had completed. The macro-script's timing was similarly softened - while previous Basilica tutors would immediately interrupt their sub-scripts for a high-level timeout, in this tutor a pending prompt for the next phase could be delayed long enough for the current sequence to finish.

TABLE 1
EXAMPLES OF REVOICING AGENT'S RESPONSES TO ACTUAL STUDENT EXPLANATIONS

Student Contribution	Revoicing Agent Response
basically the glucose will get inside	Maybe you could state that as "the cell membrane is permeable to glucose."
it changed becouse the tube obsorbed the iodine,	So are you saying "the molecules diffused through the membrane?"
I predict that if the holes in the plastic are large enough, the glucose will go into the water solution..	Maybe you could state that as "both water and glucose molecules are able to move between the two environments."

3.3 Academically Productive Feedback

Another manipulation implemented using Bazaar and evaluated in this study is an agent that provides positive feedback for APT. Student input was matched against a list of patterns indicating APT moves, including explanation, challenge, revoicing, and requests for others to provide each of the same. If a student statement matched, the agent publicly praised the student's move, and (when appropriate) encouraged the other students to respond. All students who participated in the study reported in this paper received training in APT prior to the online collaborative activity. Rather than perform APT based facilitation itself, as the Revoicing behavior did, the Feedback behavior was meant to indirectly support the prevalence of APT in the discussions by encouraging students to take this facilitation role.

4 METHOD

4.1 Instructional Content and Study Procedure

Participants: This study was conducted in 7 9th grade biology classes of an urban school district. The classes were distributed across two teachers (with respectively 3 and 4 classes) for a total of 78 consenting students.

Experimental Manipulation: This study was run as a 2x2 factorial design in which the APT agents provided different behaviours. Across all conditions, the agent provided the same macro level support by guiding the students through the activity using the same phases introduced in such as way as to control for time on task. The conditions of the study were defined based on the microscripting behaviors of the agent. The first variable for manipulation was the presence or absence of the Revoicing behavior described in Section 3.4. The second variable was the presence or absence of the APT Feedback behavior described in Section 3.5.

In addition, in each class, a group was provided with "wizard of oz" support in which a human experimenter performed both revoicing and feedback. We did this in order to assess whether any deficiency in positive effect of either factor might be due to technical failure rather than poor design. Results in the Wizard conditions on all measures were always within the same range as in the fully automatic support conditions.

Learning Content: The study was carried out during a module introducing the concepts of selective permeability, diffusion, osmosis and equilibrium. In this module, students observe that glucose, water and iodine molecules all diffuse through dialysis tubing while starch molecules do not. The activity naturally lends itself to a variety of distinct cell models involving dialysis tubing containing an *inside environment* immersed in a beaker containing the *outside environment*. In each, a choice must be made for which liquid will be placed outside and which liquid will be placed inside. Four were used in the study:

1. Model A includes a starch suspension inside dialysis tubing and iodine solution outside (the iodine serves as an indicator for starch).
2. Model B is the opposite of A, having the iodine solution within the dialysis tubing and the starch suspention outside.
3. Model C includes a glucose solution on the inside of the dialysis tubing and distilled water on the outside.
4. Model D is the opposite of C. It has distilled water in dialysis tubing and glucose solution on the outside.

In the case of cell models A and B, movement of the starch suspension and iodine solution can be detected through a change in color of the inside or outside environment. In the other two cell models, indicator strips that change color in the presence of glucose can detect whether the glucose solution has mixed with the distilled water.

Study Procedure: The study was conducted over three phases, which occurred as single class periods over two school days.

The first phase ("day 1") involved the teachers running a lab as a demonstration of building a cell model with their students as they would normally with cell model A, the condition of starch suspension inside dialysis tubing and iodine solution outside. The students observe the cell model as it is constructed and then 24 hours later. The students took a pre-test at the end of this first phase.

The second phase ("day 2") was centered around a 20mn collaborative computer-mediated activity during which the experimental manipulation took place. The students did the activity in groups of 3 students, scaffolded by Academically Productive Talk conversational agents. Students within classes were randomly assigned to groups and then groups to conditions This activity was introduced by a cartoon depicting the use of APT, a reminder of the results of the previous day (with cell model A) and an introduction to the "new" information: glucose and glucose test strips. The conversational agent led the students through two new conditions: cell models B and C. For each of these conditions, the agent showed the outcomes after 1 and 24 hours in terms of the colors inside and outside (indicating whether starch and glucose had diffused in or out) and the weight of the tubing (indicating whether water had travelled). For each observation, the agent asked the students to come up with an explanation. The agent then presented the students with cell model D, glucose outside and water inside (the opposite of model C) and asked the students to collaboratively come up with a prediction for what they would observe, and an explanation for their prediction. They were instructed to write down their prediction and explanation when they were in agreement and were informed that there would be prizes for the best explanations. To assist them in this activity, students were given a worksheet summarizing the setup for each condition and providing space to write down their prediction and explanation for cell model D. Since the students talked over the explanations before recording them on their respective sheets, we refer to these as Co-constructed explanations. At the end of this second phase, the students took the Post-Activity test.

The computer activity was intended to equip the students with enough empirical data and attempts at reasoning to prepare them for the third phase ("day 3"), a full class APT discussion with their teacher, during which they would reconcile their different understandings and explanations. At the end of this discussion, they took a second post-test.

4.2 Measurement

Domain knowledge was measured at three time points using a paper based test. Each of the three tests (pre-test, post-activity test, post-discussion-test) followed a similar format: a multiple choice question, a fill in the blank question and what we refer to as a concept cartoon. Each test was different and was specifically targeted to the kinds of knowledge the students might be expected to have at that associated phase of the activity. The idea of the concept cartoon is to present a contextualized situation with three statements which can all be true given certain assumptions. Respondents are asked to pick the statement they are most in agreement with and to explain why they agree. Each of the concept cartoon explanations were graded in a similar way along two dimensions: the number of science terms (e.g. "diffuse through the membrane" as opposed to "went through the bag") and the degree of understanding exhibited in the explanation provided. A similar rubric was also used to evaluate the predictions and explanations that students wrote at the end of the collaborative learning session prior to the post-Activity test, namely the Co-constructed explanation.

4.3 Results

In this study we have tested the hypothesis that offering dynamic microscripting support to computer supported collaborative learning groups in the style of Academically Productive Talk (APT) facilitation will produce more learning during collaborative learning discussions by enriching the interactions between students, and will also better prepare them for participation in a whole group, teacher lead discussion.

As mentioned above, two independent manipulations were used to operationalize APT facilitation in this study, namely Revoicing and Feedback. In order to evaluate the hypothesis, we took 5 measurements. First, in order to measure learning, we offered a pretest, post-activity test, and post discussion test. Learning specifically between Pre-test and Post-Activity test is learning during the experimental manipulation. Second, in order to measure the extent to which the online discussion lead to successful group collaboration, we also measured the quality of the joint product produced in the online activity, which was an explanation for a prediction that was co-constructed by the group and then written separately by each student on their worksheet. Finally, in order to measure preparation for participation in the whole group discussion, we also measured the frequency of participation in that discussion and evaluated learning between the Post-Activity test and the Post-Discussion test.

The results per condition are summarized in Table 2. We found support for the first part of this hypothesis, namely that one form of APT support, in particular Revoicing, had a significant positive effect on learning during collaborative learning as well as resulting in higher quality co-constructed explanations. However, we did not find support for the second part of the hypothesis related to preparation for learning during a whole class teacher lead discussion. In this section we detail our analyses as well as some alternative explanations for the pattern of results we have found.

TABLE 2
SUMMARY OF RESULTS PER CONDITION, MEAN (STANDARD DEVIATION)

	Control Condition	**Feedback Only**	**Revoicing Only**	**Revoicing and Feedback**
PreTest	.48 (.32)	.43 (.32)	.56 (.2)	.45 (.3)
Post-Activity Test	.4 (.3)	.34 (.31)	.52 (.24)	.46 (.23)
Post Discussion Test	.46 (.21)	.45 (.17)	.53 (.19)	.48 (.18)
Co-Constructed Explanation	.66 (.28)	.46 (.3)	.7 (.22)	.63 (.18)
Class Discussion Participation (normalized)	.01 (.01)	.03 (.04)	.03 (.04)	.02 (.02)

We began our analysis by evaluating the effect of the experimental manipulation on learning. First, we confirmed that our random assignment was successful in assigning students to groups that were roughly equivalent with respect to prior knowledge. We did this by using an ANOVA, with Revoicing and Feedback as independent variables and Pre-Test as the dependent variable. We also included an interaction term for the interaction between Revoicing and Feedback. There were no significant or marginal effects of either independent variable or the interaction.

Then we tested the effect of the experimental manipulation on learning during the collaborative activity using an ANCOVA with Post-Activity test as the dependent variable, Revoicing and Feedback as independent variables and Pre-test as a covariate. We also included the interaction term between Revoicing and Feedback. There was a significant positive effect of Revoicing $F(1,69) = 4.4$, $p < .05$, effect size .47 s.d., but no significant effect of Feedback, and no interaction between the two factors.

We also examined the effect of the experimental manipulation on the quality of the co-constructed explanation that the students came up with in their groups. We evaluated this by using an ANOVA with Co-constructed explanation score as the dependent variable and Revoicing and Feedback as independent variables. There was no significant correlation between this score and Pre-test score, so we did not use Pre-test as a covariate in this analysis. There was a significant positive effect of Revoicing $F(1,74) = 4.3$, $p < .05$, effect size .48 s.d., and a significant negative effect of Feedback $F(1,74) = 4.89$, $p < .05$, effect size .51 s.d.

Next we examined the Post-Discussion test scores, which were from the test students took after participating in the whole class, teacher lead discussion. We examined this two ways. First we examined learning between Pre-test and Post-discussion test in order to determine the effect of the online activity in combination with the whole class discussion. Then we examined learning between the Post-activity test and the Post-discussion test just to evaluate the effect of the manipulation on preparation for learning from a whole class discussion. There was no significant effect of the manipulation or the interaction between factors on the Post-Discussion test score either with Pre-test as a covariate or with Post-Activity test as a covariate. Thus, we must conclude that students learned the same amount during the whole class discussion regardless of condition, and the learning that occurred during that discussion washed out the effect of the experimental manipulation on learning during the online activity.

We also looked for evidence that the experimental manipulation may have affected participation in the whole class teacher lead discussion. In order to evaluate this, we did an ANOVA with Class Discussion Participation as the dependent variable and Revoicing and Feedback as independent variables. We also included the interaction between these two factors. There was no significant effect of either factor or the interaction.

4.4 Discussion

What we see here first is a local effect of learning during the collaborative activity, which did not result in a persistant effect on achievement. Although students in the Revoicing condition came out of the collaborative activity with a domain knowledge advantage, the other students were able to catch up during the whole class discussion, in which students from all conditions participated together, with the support of their teacher, who employed APT discussion techniques in the whole class discussion.

There are at least three possible explanations for the catch up effect, which we will investigate in our future work. One possible explanation is that students in the whole class discussion who came in with less knowledge benefitted from hearing the contributions of their more knowledgeable peers from other conditions. In order to further explore this option, in future research we may conduct whole class discussions separately for each condition in order to avoid a possible "bleed over" effect.

Another related possibility is that the fact that the teacher was using APT facilitation had a positive leveling effect on knowledge within the classroom. Alternatively, the immediate learning effect may have simply not resulted in long term retention of knowledge. It is notable that only in the condition with both Revoicing and Feedback do we see a consistent positive trend in test scores from Pre-test, to Post-Activity test, to Post-Discussion test. Since we did not counter-balance the tests that were offered to students at these distinct time points, it may be that the Pre-test was somewhat easier than the other two tests. One way of separating these effects would be to have half the class participate in the large group discussion while the other half engage in a different activity prior to the third test.

Beyond the short versus long term learning issue, another interesting finding from this study was a differential effect of the two distinct APT manipulations. Whereas Revoicing had a positive effect on learning as well as on the quality of the co-constructed explanations, Feedback had no effect on learning and a negative effect on the quality of the co-constructed explanations. Further investigation into the nature of the discussions that took place in the different conditions will be needed to understand how the manipulations lead to differing effects. The simple conclusion that may be drawn from the result indicates that using agents to perform APT based facilitation may be more effective than attempting to encourage students to play that role for each other.

5 Conclusions and Current Directions

This article presents a first successful evaluation of a new form of dynamic support for collaborative learning that was inspired by the work in the classroom discourse community on Academically Productive Talk. This form of support was implemented within a new agent based architecture called Bazaar, which extends earlier work with the Basilica architecture. The proposed dynamic support approach was evaluated in a classroom study involving 7 9^{th} grade biology classes in an urban school district. The study provides evidence that form of the support, namely the Revoicing support, lead to significantly more learning within an online collaborative learning activity and resulted in higher quality explanations for predictions within that session. Future work will investigate the reasons for the differential effect between the two forms of APT based support as well as the "catch up" effect discovered as a result of the whole group teacher lead discussion using APT that occurred in the class session on the day immediately following that of the online activity.

Acknowledgment

The authors wish to thank Lauren Resnick, Sandra Katz, Catherine Stainton, and Sherice Clarke for their involvement in the district wide teacher training effort that provides the context for this work. This work was supported in part by NSF grant SBE 0836012 to the Pittsburgh Science of Learning Center.

References

[1] Ai, H., Kumar, R., Nguyen, D., Nagasunder, A., Rosé, C. P. (2010). Exploring the Effectiveness of Social Capabilities and Goal Alignment in Computer Supported Collaborative Learning, in *Proceedings of Intelligent Tutoring Systems.*

[2] Bales, R.F. (1950). *Interaction process analysis: A method for the study of small groups*. Addison-Wesley

[3] Berkowitz, M., & Gibbs, J. (1983). Measuring the developmental features of moral discussion. *Merrill-Palmer Quarterly*, 29, pp 399-410.

[4] Dillenbourg, P. (2002). Over-scripting CSCL : The risks of blending collaborative learning with instructional design . *Three worlds of CSCL: Can we support CSCL*, pp. 61-91.

[5] Dillenbourg, P., Tchounikine, P. (2007). Flexibility in macro-scripts for computer-supported collaborative learning. *Journal of Computer Assisted Learning* 23(1), pp 1-13.

[6] Dillenbourg, P., Hong, F. (2008). The mechanics of CSCL macro scripts. *The International Journal of Computer-Supported Collaborative Learning* 3(1), pp 5-23.

[7] Fernando, S., and Stevenson, M. (2008). A semantic similarity approach to paraphrase detection, *Computational Linguistics UK (CLUK 2008) 11th Annual Research Colloquium.*

[8] Kobbe, L., Weinberger, A., Dillenbourg, P., Harrer, A., Hämäläinen, R., Häkkinen, P., Fischer, F. (2007). Specifying computer-supported collaboration scripts. *The International Journal of Computer-Supported Collaborative Learning* 2(2-3), pp 211-224.

[9] Kollar, I., Fischer, F., Hesse, F.W. (2006). Collaborative scripts - a conceptual analysis. *Educational Psychology Review* 18(2), pp 159-185.

[10] Kumar, R., Gweon, G., Joshi, M., Cui, Y., Rosé, C. P. (2007). Supporting Students Working Together on Math with Social Dialogue. *Proceedings of the SLaTE Workshop on Speech and Language Technology in Education*

[11] Kumar, R., Rosé, C. P., Wang, Y. C., Joshi, M., Robinson, A. (2007). Tutorial Dialogue as Adaptive Collaborative Learning Support, *Proceedings of Artificial Intelligence in Education.*

[12] Kumar, R., Ai, H., Beuth, J., Rosé, C. P. (2010). Socially-capable Conversational Tutors can be Effective in Collaborative Learning Situations, in *Proceedings of Intelligent Tutoring Systems*.

[13] Kumar, R., Rosé, C.P. (2011). Architecture for Building Conversational Agents that Support Collaborative Learning. *IEEE Transactions on Learning Technologies* 4(1).

[14] Lison, P. (2011). Multi-Policy Dialogue Management. In *Proceedings of the SIGDIAL 2011 Conference*, Association for Computa-

tional Linguistics, pp. 294-300

[15] Michaels, S., O'Connor, C., & Resnick, L.B. (2007). Deliberative discourse idealized and realized: Accountable talk in the classroom and in civic life. *Studies in Philosophy and Education*.

[16] Mühlpfordt, M., Wessner, M. (2005). Explicit referencing in chat supports collaborative learning. *Proceedings of the 2005 conference on Computer support for collaborative learning: the next 10 years*, pp. 460-469.

[17] Resnick, L. B., Bill, V., & Lesgold, S. (1992). Developing thinking abilities in arithmetic class. In A. Demetriou, M. Shayer, & A. Efklides (Eds.), *Neo-Piagetian theories of cognitive development: Implications and applications for education* (pp. 210-230). London: Routledge

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

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

[20] Wecker, C., Fischer, F. (2007). Fading scripts in computer-supported collaborative learning: The role of distributed monitoring. *Proceedings of the 8th iternational conference on Computer Supported Collaborative Learning*, pp. 764-772.

[21] Weinberger, A., Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education* 46(1), pp 71-95.

[22] Weinberger, A., Stegmann, K., & Fischer, F. (2007). Scripting argumentative knowledge construction: Effects on individual and collaborative learning. In C. Chinn, G. Erkens, & S. Puntambekar (Eds.), *Mice, minds, and society: CSCL 2007* (pp. 37-39). New Brunswick, NJ: International Society of the Learning Sciences.

Gregory Dyke obtained his PhD in 2009 at the Ecole des Mines de Saint-Etienne. He is currently a post-doctoral reasearcher at Carnegie Mellon University. His doctoral research into frameworks for modeling and capitalizing on analyses of interaction resulted in a student best paper award at CSCL 2009. He is interested in analytic techniques for studying collaborative learning and interaction. His current focus is on visualizations to discover temporal patterns.

David S. Adamson earned his BA in Computer Science from Oberlin College in 2004, after which he taught high school mathematics and computer science in Baltimore City from 2004 until 2010, and is presently a Master's student at the Language Technologies Institute.

Iris Howley is pursuing her PhD at Carnegie Mellon University's Human-Computer Interaction Institute. She received her Bachelor of Science from Drexel University where her research centered around applications leveraging the Semantic Web. Currently, her research spans the fields of dialogue agents, behavioral psychology, and experimental design. Her most recent projects focus on developing computational support to enable students to overcome obstacles to effective participation in collaborative learning situations.

Carolyn P. Rosé is an Associate Professor with a joint appointment between the Language Technologies Institute and the Human-Computer Interaction Institute at Carnegie Mellon University. She completed her Master's degree in Computational Linguistics from Carnegie Mellon University in 1994 and her PhD in Language and Information Technologies from Carnegie Mellon University in 1998. She has over 125 peer reviewed publications, mainly in top tier conferences and journals, spanning the fields of language technologies, learning sciences, and human-computer interaction. She serves as the Secretary/Treasurer of the International Society of the Learning Sciences and Associate Editor of the International Journal of Computer Supported Collaborative Learning.