Intensification of Group Knowledge Exchange with Academically Productive Talk Agents

David Adamson, Colin Ashe, Hyeju Jang, David Yaron, Carolyn P. Rosé
Carnegie Mellon University, Pittsburgh PA
Email: {dadamson, cashe, hyejuj, yaron, cprose}@cmu.edu

Abstract: In recent years, intelligent conversational agents have been used with some level of effectiveness as dynamic support for collaborative learning in online chat. The classroom discourse community offers insights from analysis of effective classroom discussion facilitation practices that might productively inspire the design of such facilitator agents. In this paper, we evaluate one such conversational agent-as-facilitator design, drawn from the literature on what has been termed Academically Productive Talk. Specifically we evaluate the effect of a facilitation strategy referred to as Agree/Disagree, where students are prompted to evaluate the assertions of a partner student. In a simple two condition study, we evaluate the effect of this facilitation strategy in comparison with an otherwise identical condition where this facilitation strategy is absent. The results demonstrate a marginal positive effect on learning (effect size .55 standard deviations) and a significant intensification effect on the collaborative discourse.

Introduction

The literature on scripted support for Computer Supported Collaborative Learning describes scripts as a set of scaffolds and interventions that structure and facilitate student interaction, at both the macro-level of the collaborative activity and at the micro-level of individual actions (Dillenbourg, 2008). In particular, an instructor's role is to orchestrate multiple scripts (Fischer & Dillenbourg, 2006) to provide comprehensive, suitable support for the students throughout the collaborative learning experience. Recently, work building on this body of research has explored the role of dynamically scripted support for CSCL in the form of conversational agents, which have been shown to be successful in promoting student learning and conversation in collaborative discussion environments (Kumar et al., 2007; Chaudhuri et al., 2009; Dyke et al., 2012).

Additionally, analyses of expert teacher talk (Chapin et al., 2003) have revealed a set of discursive instructional practices, suitable for facilitating collaborative knowledge-building. The Academically Productive Talk framework (Michaels et al., 2007) describes a collection of discussion-facilitating moves a teacher can be employed to promote rich student-centered conversation and collaboration. This framework can serve as an operationalization of effective group facilitation techniques which, combined with results and experiences from the CSCL scripting and conversational agents communities, lays the groundwork for automatic agent-based facilitation of small group online chat. Recent studies have made important advances in this area, and have identified limitations in agent design and behavior that must still be overcome. The contribution of this paper is to describe a successful new conversational agent behavior based on the principles of Academically Productive Talk, whose use leads to demonstrable gains in conceptually-rich student conversation and shows promising results for student learning.

In the remainder of the paper we first briefly review the literature on Academically Productive Talk and how it motivates design of intelligent conversational agent based support for collaborative learning as a form of dynamic microscripting. Next we describe our experimental design and methodology for process analysis. Then we describe our results and offer some interpretation. We conclude with a discussion of some limitations of this work and our current research directions.

Theoretical Background

The work presented here builds upon prior work from two disciplines: the discursive instructional framework of *Academically Productive Talk,* and the extensive body of CSCL research on supporting collaboration through scripting and conversational agents.

Academically Productive Talk

Academically Productive Talk has grown out of frameworks that emphasize the importance of social interaction in the development of mental processes. Michaels, O'Connor and Resnick (Michaels et al., 2007) describe a number of core moves that discussion facilitators can employ to foster effective student-centered classroom discussion. A selection of these moves are presented in Table 1.

Table 1. Selected Accountable Talk Moves

Academically Productive Talk Move	Example
Revoicing a student's statement	"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)
Asking students to *restate* someone else's reasoning	"Can you repeat what she just said, in your own words?"
Asking students to apply their own reasoning to someone else's reasoning	"Do you *agree or disagree*, and why?"

The teacher's facilitation plays a key role in encouraging transactive conversational behavior between students, but, importantly, does not lead to a teacher-centered discussion. Instead, the teacher uses Academically Productive Talk to hold students accountable for their own knowledge and reasoning, and to remind them to hold themselves and each other accountable likewise. In studies where teachers used approaches like Academically Productive Talk, students have shown steep changes in achievement on standardized math scores, transfer to reading test scores, and retention of transfer for up to 3 years (Bill et al., 1992; Chapin et al., 2004). In another recent study, urban high-school teachers were trained in Academically Productive Talk practices. During the same period, the teachers' students participated in computer-supported collaborative learning activities that promoted Academically Productive Talk. Over the course of the study and especially following the interventions, the amount of Academically Productive Talk moves performed in the classroom was shown to increase (Clarke et al., this volume)

Script-Based Support for Collaborative Learning

The CSCL community shares many of the same values related to desired conversational practices in student group discussions. To support the growth of student discussion skills, we can design environments with affordances that play the same role as the teacher-as-discussion-facilitator.

The most popular approach to providing such affordances in the past decade has been that of script-based collaboration (Dillenbourg, 2002). A script may provide structure at a macro-level, perhaps dividing a collaborative task into roles for the participants to fulfill, or might scaffold a participant's contributions at a micro-level, with prompts to encourage a particular mode of argumentation. Such scripts are typically implemented statically, providing the same support in all cases. This is the work we review in this section. In the next section we describe a dynamic form of scripting that is capable of responding to changes in the state of the environment or discussion to deliver an appropriate level of support at opportune times.

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. Scripts can be classified as either macro-scripts or micro-scripts (Dillenbourg, 2008). Macro-scripts are pedagogical models that describe coarse-grained features of a collaborative setting, that sequence and structure each phase of a group's activities to foster learning and social interaction. 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. Scripts can be more or less coercive, from strict "follow me" style prompts to subtle suggestions of behavior implicit in the activity's structure. Stricter scripts can work to reduce the gap between expected and observed student behavior, producing a more uniform appearance of discussion, but run the risk over-scripting (Dillenbourg, 2002), where the application of inappropriate or unneeded supports have a detrimental effect on collaboration and learning.

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, or to the interference between multiple scripts (Weinberger et al., 2007). More dynamic approaches can trigger scripted support in response to the automatic analysis of participant activity (Rosé et al., 2008). This analysis can occur at a macro-level, following the state of the activity as a whole, or it could be based on the micro-level classification of individual user contributions. The collaborative tutoring agents described by (Kumar & Rosé, 2011) were among the first to implement dynamic scripting in a CSCL environment. Scripting such as this offers the potential for minimal interventions to be used more precisely and to greater effect, with greater likelihood of students internalizing the support's intended interaction patterns. Further, the benefits of fading support over time (Wecker & Fischer, 2007) could be more fully realized, as the frequency of intervention could be tuned to the students' demonstrated competence. Indeed, conversational agents have been shown to be more effective when their interaction with students is in response to student initiative (Chaudhuri et al., 2009).

Participants in a collaborative session, including the facilitator, aren't simply focused on the task – they are involved in numerous simultaneous processes including social bonding, idea formation, argumentation, time management, and off-task activity. Just as human teachers orchestrate elements of collaborative learning in their classrooms, a conversational agent-as-facilitator must manage several differently-scoped supports and behaviors

concurrently. Recent work has produced software architectures for conversational agents (Kumar & Rosé, 2011; Adamson et al., 2012) that can implement such orchestration within CSCL environments.

Agents for Academically Productive Talk

Prior work with conversational agents and Academically Productive Talk has directed students to respond to each other with an array of Academically Productive Talk moves, in response to surface-level features of their contributions, with mixed results, prompting a redesign wherein the agent offered "Revoice" prompts that paraphrased student contributions when they were identified as conceptually-rich and relevant to the task (Dyke et al., 2012). Such an agent was shown to have a positive effect on learning and on conceptual richness of later student contributions. Criticism of the agents used in these studies (Stahl, 2013) suggests that the student experience could be improved by more finely targeting its interventions such that they are more responsive to (and not disruptive of) the flow of collaboration, and by minimizing the verbosity of each agent contribution.

We present a conversational agent behavior based on the "Agree-Disagree" Academically Productive Talk move as a dynamic support within a scripted CSCL environment, addressing some the limitations found in earlier work. In our implementation, the conversational agent acts as an instructor and facilitator, and presents a series of group exercises in ConcertChat, a discussion environment with a shared whiteboard (Mühlpfordt & Wessner, 2005). This environment is illustrated in Figure 1. As the group discusses each exercise, the agent monitors the chat for student assertions that could be followed up by a check for agreement or understanding. After such a candidate is identified, the agent waits to see if the students address the assertion on their own – if not, the agent offers a prompt to focus the group on the student's contribution.

Figure 1. Screen shot of the CSCL environment where a group of 3 students is working together, supported by a tutor agent named Quinn, who participates with them in the chat.

Detecting Academically Productive Talk Candidates

In order to identify task-relevant conceptual assertions, we worked with domain experts and instructors to develop a "gold standard" list of statements that captured important concepts and misconceptions for the unit of study. Such statements were drawn from both the experts' knowledge and expectations and from transcripts of an unsupported dry-run of the task. Using a "bag of synonyms" cosine similarity measure (Mihalcea et al., 2006), which essentially measures overlap in word usage, student assertions which are within a certain threshold of similarity to the gold statements are identified as *agree-disagree candidates* that could be evaluated by the group. This is the same detection technique used by the earlier Revoicing agent behavior (Dyke, et al. 2012), although as the agent does not need to produce an accurate paraphrase from the matched statements, a lower threshold can be used. This results in the detection of a greater number of candidate statements, and more

opportunities for support than the Revoicing agent could afford. Statements that match only the stricter threshold are also tracked – these *revoicable* assertions serve as a conservative indicator of conceptual, on-target contributions by each student. In earlier studies, the number of revoicable assertions was found to significantly correlate with learning. In this study, we expected the Agree-Disagree agent to intensify the contribution of this type of valued contribution by students.

Responding to Candidates

When a candidate statement is identified, the agent waits for the other students in the group to respond to it. If another student responds with an evaluation of their peer's contribution (along the lines of "I agree" or "I think you're wrong"), but doesn't support the evaluation with an explanation, the agent will encourage this second student to provide one. If a student instead follows up with another APT candidate statement, the agent does nothing, leaving the floor open for productive student discussion to continue unimpeded, reducing the risk of over-scripting their collaboration. If the other students do not respond with either an evaluation or a contentful followup, the agent prompts them to comment on the candidate statement – for example, "What do you think about Student's idea? Do you agree or disagree?" This process is illustrated in Figure 2.
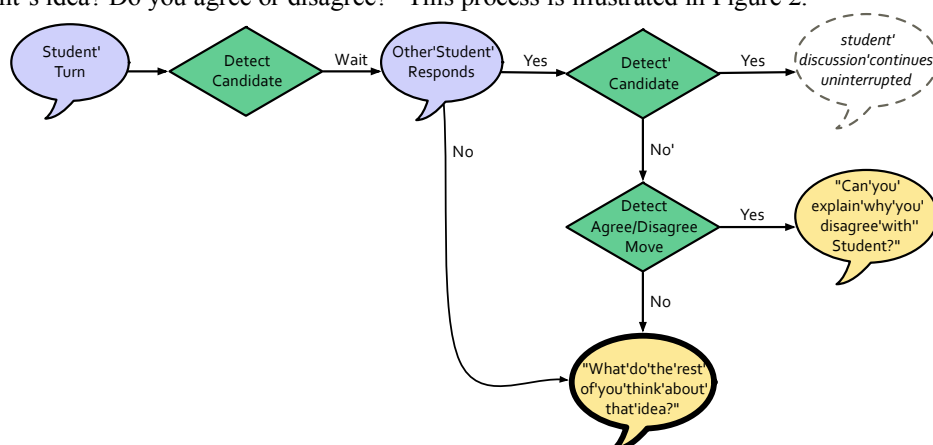

Figure 2. the Agree-Disagree Agent's response to student statements

The excerpt shown in Table 2 is drawn from the study described in the next section. Times are given in seconds from the beginning of the excerpt, and the columns "Agree/Disagree Candidate" and "Student Evaluation" are the automatically detected labels the agent uses to motivate its facilitation moves. This exchange is typical of a group interaction in this environment – S07's contribution at 17 seconds, although a candidate for evaluation, is not acted upon by the agent because S08's followup preempts it. After 15 seconds following S08's statement without any sort of uptake by the group, the agent prompts the group to agree or disagree with it. S09 offers a challenge, which leads to an extended back-and-forth between all three students.

Table 2: Selected interaction with the Agree-Disagree agent

Time	Author	Text	Agree/Disagree Candidate	Student Evaluation
00:00	S07	ok lots of things to do...		
00:13	S07	first one		
00:17	S07	surface area is higher	✓	
00:20	S09	arrow up?		
00:22	S08	ok, boiling pt will go up and vdW will go up for all of them consecutively... right?	✓	
00:37	TUTOR	Do you concur with S08? Why, or why not?		
00:41	S09	**hmm not necessarily**		✓
00:47	S07	area goes up for each	✓	
00:50	S09	would it?		
00:51	S09	**im not sure**		✓
00:56	S08	yea for sure area goes up	✓	
01:10	S07	dipole increases first one		

While this approach goes far in providing productive prompts at appropriate points, students can still be thrown off by these interventions. In Table 3, the agent does not identify the ongoing exchange as relevant to the discussion, and thus does not suppress its prompt for evaluating S08's earlier statement. This causes

confusion for S08, who is unclear about which of their messages the agent is referring to. Occasional missteps such as this do not appear to utterly derail the group and, the agent is generally accepted as a facilitator and its prompts are taken as opportunities for reflection.

Table 3: Infelicitous interaction with the Agree-Disagree agent

Time	Author	Text	Agree/Disagree Candidate	Student Evaluation
01:10	S07	dipole increases first one		
01:13	S08	dipole moment is based on the whole thing though, and it's tetrahedral... agh	✓	
01:14	S07	then its symmetric?		
01:16	S08	shapes are hard		
01:19	S07	so decrease		
01:24	S07	and then increase		
01:27	TUTOR	What do you think about S08's idea? Do you agree or disagree?		
01:27	S07	and then decrease?		
01:29	S08	wait what?		
01:49	S08	TUTOR, if it ends in a question mark its probs not an idea		✓
01:54	S07	CF4 is symmetric so dipole would be 0?	✓	

Other Agent Behaviors

We employ the Bazaar agent framework (Adamson et al., 2012) to dynamically orchestrate the full set of agent behaviors, prioritizing and regulating the proposed contributions from each of the agent's components proposed contributions, so as to avoid interference between components, and to work with the flow of the group's conversation. In addition to the agree-disagree behavior described above, the agent executes a flexibly-timed macro-script to present a series of instructional materials and exercises on the group's shared whiteboard. This script begins when a sufficient number of students have joined the group chat. When all students indicate that they are ready to proceed to the next phase of the task, the agent clears the whiteboard and presents the material for the next problem. The agent also implements a set of social support moves, providing responses to student behavioral cues in order to promote group bonding and task-oriented positivity. Such support has been shown to correlate with gains in student learning and perception of the agent (Kumar et al., 2010, Ai et al., 2011).

Method

To investigate the efficacy of the Agree-Disagree agent as a way to promote student interaction and critical thinking, we situated our study within a first-year undergraduate chemistry course.

Participants

The participants in our study were first-year undergraduate students studying intermolecular forces. Students were randomly assigned to groups of 3 or 4, and then groups were randomly assigned to conditions. The balance of 3 and 4 person groups was even between conditions, and there was no effect of team size on any of our dependent measures. All students in the course were required to participate in the online exercise for course credit, but they had the option of not consenting for their data to be included in our research. Thus, we only report results for consenting students. Altogether, our analysis includes data from 18 students from 6 different groups, which is 9 students and 3 groups in each condition. We employ multi-level modeling techniques in our analyses of results in order to account for the statistical dependencies between data from students in the same group.

Task

The collaborative task focused on intermolecular forces and their influence on the boiling points of liquids. For each problem in the activity (illustrated in Figure 1), students were asked to predict whether a given substance would have a higher or lower boiling point than two of its relatives, explaining their reasoning about the set of molecules in terms of their structure and the forces at play. Each problem of this sort was followed up by revealing the actual boiling point of the mystery molecule, and asking students to revisit their predictions and explanations in light of the new data. A liquid's boiling point can be influenced simultaneously by a number of different intermolecular forces, each of which arises as a consequence of the molecules' particular structural attributes. Correctly identifying the pertinent structural features of molecules and reasoning about how they will

affect the liquid's boiling point is a non-trivial and multi-faceted task. Because multiple types of intermolecular forces influence liquids' boiling points, we used the Jigsaw technique (Aronson, Blaney, Stephan, Sikes, & Snapp, 1978), assigning students within each group to read individually about one of three forces that contribute to a molecule's boiling point. In cases where a four-person group was formed, the fourth student received the same training material as the first student. This division also provided intrinsic motivation for collaboration, as the task could not be completed without knowledge from each of the student experts.

Experimental Design

Our experimental design was a simple 2-condition between-subjects design where teams were assigned randomly either to the Agree-Disagree condition or the Control condition. Both conditions were identical except for inclusion of the Agree-Disagree facilitation move by the agent. Thus, both conditions benefitted both from macro-level and micro-level script based support. In the Agree-Disagree condition, whenever the agent was not engaged in a directed dialog, it was receptive to opportunities to dynamically support the conversation by requesting students to evaluate whether they agreed or disagreed with assertions that were made in the chat, as discussed above.

Pre/Post Tests

Pre and Post tests were used to measure learning during the collaborative exercise. We used two isomorphic versions of the test (Version A and Version B) and counter-balanced their assignment such that half of the students received A as a pretest and B as a posttest, while the other half of students received B as pretest and A as posttest. There was no significant difference between scores on A and B.

Process Analysis

The goal of the Agree/Disagree agent was to engage students in a more intensive exchange of explanations (revoicable assertions), to raise the level of critical thinking. Thus, in addition to a Pre/Post test measure of learning, a process analysis is also important for evaluating our hypothesis. Variables related to the elicited conversational behavior may then be examined in order to test whether they served a mediating or moderating effect on learning. In order to accomplish this, the chat logs were segmented into 2 minute intervals such that one observation was extracted per student for each interval. In each observation, we counted the number of revoicable assertions contributed by the student, the number of revoicable assertions contributed by other group members, the number of Agree-Disagree prompts targeted at the student in the previous time slice, and the number of Agree-Disagree prompts targeted at other students in the group in the previous time slice.

We can evaluate the effect of condition on the correlation within time slices between occurrences of revoicable assertions of a student with those of the other students in the same group. We used a multi-level model to analyze the results in order to account for group effects. We expect to see that the correlation is significantly higher in the condition with the Agree/Disagree agent. Specifically, we used what is referred to as a *random intercept and slope model*, which allows estimating a separate latent trajectory for a student's behavior in relation to that of their partner students within time slices. In this model, each student trajectory is characterized by a regression with latent slope and intercept, relative to a slope and intercept per group, which are in turn relative to the global model's slope and intercept. To do this analysis, we used the Generalized Linear Latent and Mixed Models (GLLAMM) (Rabe-Hesketh, Skrondal, & Pickles, 2004) add-on to STATA (Rabe-Hesketh & Skrondal, 2012). The dependent measure was number of revoicable assertions by the student within the time slice. The independent variable was the number of revoicable assertions contributed by the other students in the group within the same time slice. The condition variable was added as a fixed effect, and as an interaction term with the independent variable. A significant interaction between condition and independent variable in this case would indicate a significant difference in correlation between a student's contribution of revoicable assertions and that of their partner students.

Results/Analysis

Our hypothesis was that the introduction of the Agree/Disagree agent would intensify the interaction between students, which might increase critical thinking, and subsequently increase learning. Our analysis offers qualified support for the hypothesis.

First we evaluated the effect of condition on learning. For this analysis, we tested for any significant difference in pretest scores between conditions using an ANOVA with pretest as a dependent variable, Condition as an independent variable, and Group as a random variable nested within condition in order to account for the non-independence between data collected from students who worked in the same group. There was no significant or marginal effect of Group on pretest scores, confirming that students were distributed with sufficient randomness between groups. There was no significant or marginal difference between conditions on pretest score, though there was a trend for students in the Agree/Disagree condition to have lower pretest scores.

Table 4: Summary of Results

	Agree/Disagree Condition	Control Condition
Pretest	6.14 (3.7)	6.73 (3.1)
Posttest	9.99 (2.72)	8.81 (3.3)
Revocable Assertions	7.89 (3.8)	7.33 (4.3)

Thus to evaluate the effect of condition on learning, we used an ANCOVA with posttest as the dependent variable, pretest as a covariate, Condition as an independent variable, and Group nested within condition as a random variable. In this analysis, there was a marginal effect of Condition on learning ($F(1,11)$ = 1.82, $p < .1$, effect size .55 standard deviations), such that students in the Agree/Disagree condition learned more. The effect was moderate.

Next we examined the intensifying effect of the intervention on the interaction between students. We evaluated this by looking for evidence that the Agree/Disagree prompts increased the extent to which students constructed knowledge together, at least in pockets of intensive knowledge exchange. As can be seen in the conversation excerpts above, students contribute a variety of types of contributions, not all of which are revocable assertions. However, when they are engaged in intensive exchange of ideas with one another, we find regions of the conversation with denser concentrations of revocable assertions, because when one student offers his perspective, others tend to follow up with their own. If the discussion is divided into time slices, we can distinguish transcripts that contain regions of dense group knowledge construction from those where students present their ideas intermittently without precipitating intensive group knowledge construction. We do this by looking at the correlation of the count of a student's revocable assertions with the count of revocable assertions from other students in the group, within each time slice. In the first case, we expect that there will be many time slices where there are revocable assertions from both the student and the other students in the group, whereas in the second condition, we don't expect to see this occur frequently.

The analysis using the random intercept and slope model described in the Methods section showed the pattern that we expected. There was no significant difference in intercept between conditions, confirming that, as we suspect from Table 4, there was no difference in absolute number of revocable assertions between conditions. However, this is not problematic since the number of revocable assertions was found to have a moderating but not mediating effect on learning. Specifically, when the revocable assertions variable was added to the ANCOVA evaluating the effect of Condition on learning as an additional covariate, it had a significant positive correlation with posttest score that increased the percent of posttest variance explained from 69% to 83% but did not reduce the effect of Condition on learning. Thus, we must conclude that the effect of condition on learning is not explainable by this simple summative measure.

More importantly, there was no significant correlation between the number of revocable assertions of a student and that of his partner students in the control condition where there was not an Agree/Disagree agent. However, there was a significant interaction between the condition variable and the number of revocable assertions contributed by partner students ($R = .14$, $z = 2.03$, $p < .05$), indicating that in the Agree/Disagree condition, there was a significant positive correlation between the number of revocable assertions contributed by a student and that contributed by partner students. Thus, we do see evidence that the intervention had the effect of precipitating pockets of intensive discussion.

We then evaluated the extent to which this effect was explained by the local presence of Agree/Disagree prompts. Surprisingly, a student contributes significantly more revocable assertions in time slices following ones wherein the agent prompted the other students to agree or disagree with that student ($F(1,847) = 4.9$, $p < .05$, effect size .35 standard deviations) but not when the agent asked the group to agree or disagree with a different student (no significant effect). And time slices with revocable assertions from both students were not primarily the same ones that contained prompts for Agree/Disagree. This suggests that the primary, or at least first, effect of the prompt may in fact be to elicit followup explanations from a student rather than to elicit feedback from the other students. Seen in conjunction with the correlation analysis above, it is possible that the prompts more often first elicited followup explanation from the student who contributed the initial agree-disagree candidate, and in response to this elaboration, the other students were drawn in and responded in turn. Thus, we see a subtle ripple effect of the intervention that is not easily quantified, even in the analysis of intensification above.

Discussion

We have described and demonstrated the effectiveness of a new conversational agent behavior in a college chemistry context. Advances in its design that address sensitivity to the flow and content of student conversation differentiate this agent from similar agents in earlier work, allowing the facilitative behavior to be minimally intrusive while still actively promoting rich student-centered discussion. Future work with larger samples should provide clarification and amplification of the positive learning trend seen here. We look forward to future studies where conversational agents successfully orchestrate multiple strategies drawn from Academically Productive Talk and other instructional discourse frameworks, to provide many-dimensioned support for group

collaboration and productive discussion. Such agents may be critical in fostering effective conversation in the rapidly growing domain of distributed-learning university courses.

References

Adamson, D., & Rosé, C. (2012). Coordinating multi-dimensional support in collaborative conversational agents. In *Proceedings of Intelligent Tutoring Systems* (pp. 346-351). Springer Berlin/Heidelberg.

Aronson, E., Blaney, N., Stephan, C., Sikes, J., & Snapp, M. (1978). *The jigsaw classroom*. Sage Publications.

Bill, V. L., Leer, M. N., Reams, L. E., & Resnick, L. B. (1992). From cupcakes to equations: The structure of discourse in a primary mathematics classroom. *Verbum*, 1, 2, 63-85.

Chapin, S., O'Connor, C., & Anderson, N. (2003). *Classroom Discussions. Using Math Talk to Help Students Learn, Grades 1, 6.*

Chaudhuri, S., Kumar, R., Howley, I., & Rosé, C. P. (2009, July). Engaging collaborative learners with helping agents. In *Proceedings of the 2009 conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling* (pp. 365-372). IOS Press.

Clarke, S., Chen, G., Stainton, C., Katz, S., Greeno, J.G., Resnick, L.B., Howley, I., Adamson, D. and Rosé, C.P. (2013). The Impact of CSCL beyond the Online Environment. *Proceedings of the 10th International Conference on Computer Supported Collaborative Learning*, July 2013.

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 .

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

Dyke, G., Howley, I., Adamson, D., Rosé, C.P. (2012). Towards Academically Productive Talk Supported by Conversational Agents. *Proceedings of Intelligent Tutoring Systems 2012*.

Howley, I., Adamson, D., Dyke, G., Mayfield, E., Beuth, J., Rosé, C.P. (2012). Group Composition and Intelligent Dialogue Tutors for Impacting Students Academic Self-Efficacy. *Proceedings of Intelligent Tutoring Systems 2012.*

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

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

Kumar, R., Rosé, C. P., Wang, Y. C., Joshi, M., and Robinson, A. (2007). Tutorial dialogue as adaptive collaborative learning support. *Proceedings of Artificial Intelligence in Education*.

Kumar, R., Rosé, C.P. (2011). Architecture for Building Conversational Agents that Support Collaborative Learning. *IEEE Transactions on Learning Technologies* 4(1), 21-34.

Mihalcea, R., Corley, C., and Strapparava, C. (2006). Corpus-based and knowledge-based measures of text semantic similarity, *Proceedings of the National Conference on Artificial Intelligence (AAAI 2006),* Boston, Massachusetts, pp. 775-780.

Mühlpfordt, M., Wessner, M. (2005). Explicit referencing in chat supports collaborative learning, *Proceedings of Computer Support for Collaborative Learning (CSCL).*

Rabe-Hesketh, S. & Skrondal, A. (2012). *Multilevel and Longitudinal Modeling Using Stata*, Stata Press.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). *GLLAMM Manual.* University of California, Berkely. U. C. Berkeley Division of Biostatistics Working Paper Series, Paper 160.

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

Stahl, G. (2013). Interaction analysis of a biology chat. In D. Suthers, K. Lund, C. P. Rosé & N. Law (Eds.), *Productive multivocality in the analysis of group interactions*. New York, NY: Springer. Web: http://GerryStahl.net/pub/multivocal.pdf.

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

Weinberger, A., Stegmann, K., Fischer, F., Mandl, H. (2007). Scripting argumentative knowledge construction in computer-supported learning environments. In *Scripting Computer-Supported Collaborative Learning*, CSCL Book Series Volume 6, Chapter 6, pp191–211 .

Acknowledgements