Activity Design Models to Support the Development of High Quality Collaborative Processes in Online Settings

Marcela Borge, Pennsylvania State University, mborge@psu.edu
Yann Shiou Ong, Pennsylvania State University, yuo103@psu.edu
Carolyn Penstein Rosé, Carnegie Mellon University, cprose@cs.cmu.edu

Abstract: In this paper we assess the utility of an activity design model and different reflective activities for improving the quality of collaborative processes. Thirty-seven online students, belonging to one of 13 teams, formed the participants of the study. Teams completed five discussion sessions as part of required course activity, using one of two reflective conditions. Each team also received feedback on their performance. We assessed the quality of processes between groups using content analysis techniques. Team process measures at the first time point were used to identify groups' initial strengths and weaknesses. To assess the utility of the model and reflective assessment designs, we used a 2x5 mixed factorial design, with Condition (two levels) as a between subjects factor and Time (5 levels) as a within subjects factor. We found that students were weakest at presenting and discussing claims and both Condition and Time are significant predictors of collaborative process quality.

Keywords: sociometacognition, regulation of collaborative processes, assessment, discussion quality, online learning, online collaboration

Introduction

Collaborative learning can provide many benefits for learners, but the outcomes of collaborative activities are dependent upon the quality of collaborative interactions that occur during activity (Barron, 2003; Kozlowski & Ilgen, 2006; Stahl, 2006). This is problematic since many students lack the cognitive skills necessary to engage in high quality collaborative interactions (Borge & Carroll, 2010; Carroll, Jiang, & Borge, 2014; Fischer et al., 2013; Hogan, 1999a; Stegmann, Mu, Gehlen-Baum, & Fischer, 2011). When left to their own devices, students commonly develop dysfunctional group processes that negatively affect collective cognitive processes, learning, and performance outcomes (Barron, 2003; Borge & Carroll, 2014; Hogan 1999b; Webb & Palincsar, 1999). Thus, there is a need to develop activity design models to help students' understand what high quality collaborative processes entail and apply this knowledge to regulate existing collaborative activity.

We are currently working on the design of a CSCL environment that supports students as they learn to collectively regulate collaborative discussions. As we design the system, we are iteratively collecting data to inform the design of activities and features in the system. In this paper, we evaluate a model for the design of activities and reflective features to improve the quality of collaborative processes that occur as students discuss and think about course concepts as part of regular course activity. Our approach takes advantage of the affordances of online technology to turn the process of learning into the product to be assessed. We present findings from an implementation that took place as part of an online introductory course on information sciences and technologies, identify specific needs of the population, and examine the extent to which our activity design model and different types of reflective activities were associated with improvement of collaborative processes over time.

Related literature

Researchers have developed different types of designs to structure and guide collaborative activity, but deciding on the type and level of support can be a challenge. For example, collaborative scripts are pedagogical or dialogical models designed to optimize collaborative processes by structuring, prompting, or constraining different behaviors (Dillenbourg & Hong, 2008). However, there are trade-offs associated with scripting. On the one hand, high levels of scripting may help students produce better quality discussions and products in the short term. On the other hand, too much scripting may prevent students from recognizing gaps in learning and putting forth the cognitive efforts to fill those gaps (Bjork, 1994; Kapur & Rummel, 2009). The challenge when providing cognitive scaffolds is to ensure that they are temporary support structures and not permanent fixtures. Providing too much guidance during collaborative discourse may not be the best approach to helping students internalize sophisticated reasoning processes because an external source is telling students what to do rather than helping students learn how to regulate learning activity for themselves.

Researchers in the field of self-regulated learning have argued that effective regulation is a cyclical process of planning, doing, and reflecting (Zimmerman, 2002). To effectively self-regulate, students need to understand and think about the task in order to plan their approach, monitor themselves as they perform the task, and then reflect on their performance based on their understanding of the task in order to revise their approach (Boekarts, 1996; White & Frederiksen, 1998). Unfortunately, most students do not display optimal self-regulatory behaviors (Winne & Hadwin, 1998). Similarly, students also do not display optimal collective regulatory behaviors (Jarvela & Hadwinn, 2013). However, laboratory studies in social psychology have shown that with training teams can improve task performance through collective regulation (DeShon et al., 2004). Whether similar types of training in classroom settings could improve students' ability to self-assess and regulate collaborative processes remains to be seen.

Given the complex and evolving nature of collaborative activity, it is likely that development of sociometacognition, the executive control of collective cognitive processes, may be a necessary prerequisite to ensure that students can consistently display collaborative processes known to promote learning and problem-solving success (Barron, 2003; Jarvela & Hadwin, 2013). The ability to regulate individual cognitive activity is positively associated with higher quality individual cognitive processes and learning outcomes (Schraw, 1998; Schoenfeld, 1998; White & Frederiksen, 1998). Therefore, it is likely that the ability to regulate collective processes will facilitate higher quality collaborative activity over time.

Based on problems associated with collaborative activities and the potential benefits associated with the development of sociometacognition, we propose an activity design model for improving the quality of collaborative discussions in classroom settings that involves iterative cycles of sociometacognitive development: students prepare for collaborative discussions, engage in collaborative discussions, assess the quality of their collaborative discussions using reflective assessments (Shimoda, White, Borge, & Frederiksen, 2013), receive feedback on group process, and then repeat the cycle. The main aim of the collaborative discussions is not to engage in formal scientific inquiry, but rather to engage in questioning and deeper analysis of scientific course content. During collaborative discussions, students receive general instructions to help structure the discussion task and ordering of discussion topics, but they receive no support during the discussion session to help them select, organize, share, or interpret information, or modify their interactions. After engaging in an online discussion, students evaluate their discussion as an object of thought. By making the discussion the product to be submitted and evaluated, students have more time to think about their collective processes than they would if they had to submit another deliverable. Finally, to account for problems associated with inaccuracies of self-assessment (Burser, Larrik, Clayman, 2006; Kruger & Dunning, 1999), students calibrate their individual reflective assessments through collective discussion combined with expert feedback. Students repeat these activities for every new discussion with the aim of improving the quality of their collaborative discussions.

Currently we require students to focus on developing two capacities, information synthesis and knowledge negotiation. These two capacities are crucial to collective cognitive processes in computer supported collaborative learning (CSCL) and collaborative work (CSCW) environments (Borge & Carroll, 2010; Stahl et al., 2006; Carroll et al., 2014). Thus, they serve as the starting point in our design model to support the development of sociometacognition.

Study aims and research questions

In this study we examine the quality of collaborative processes of students enrolled in an online course and the extent to which the use of our activity design model helps students improve two core collaborative capacities: information synthesis and knowledge negotiation. Our research questions are as follows:

(RQ1) What are the most common problems students face that interfere with high quality collaborative reasoning?

(RQ2) To what extent does our activity design model and differing reflective-assessment activities affect teams' improvement of collaborative reasoning over time?

Methods

Study context

The study took place in a 16-week university level introductory online course on information sciences and technology. The main aim of the course was to introduce students to concepts and research areas central to information sciences, i.e., security and risk analysis, human computer interaction, emerging technologies, effects of technology on society, and informatics. The course was organized in a learning management system (LMS), with weekly lessons, student resources, course communication, and course materials all housed in the LMS. The instructor of the course was expected to organize and maintain the course, revise instructional materials as needed, grade student work, answer student questions, and help students to think more deeply about course content. As part of the course, students had to learn to work as part of effective teams and had to complete a team project. For this reason, developing better collaborative reasoning practices is a required part of the course and the discussion activities count towards 25% of their total grade.

Research design

We used quantitative and qualitative analysis techniques to examine students' collaborative processes and learning over time, with teams as the unit of analysis. We used a 2 x 5 mixed factorial design, with Condition (two levels) as a between subjects factor and Time (5 levels) as a within subjects factor. The quality of processes between groups was assessed using content analysis techniques. Team process measures at the first time point were used to identify groups' initial strengths and weaknesses.

Participants

Thirty-seven online students formed the participants of the study, each belonging to one of 13 groups. Eleven students (30.5%) were female and 25 students (69.4%) were male. The female to male ratio was fairly representative of the enrollment of information sciences and technology courses at this college. The groups were formed with consideration to availability for online group meetings, gender, expertise in information sciences and technology, and employment status. Groups were assigned to condition A, future thinking, or condition B, evidence, such that the groups in each condition were comparable. There were five females in condition A, six in condition B. Seventy-one percent of participants in condition A were in the 25 - 44 age range; 75% were in the 25 - 44 age range in condition B. With regard to group composition, there were five teams of three and one team of two in condition A; six teams of three and one team of two in condition B. In condition A, 66.7% of the teams were majority male compared to 71.4% in condition B. Neither group had all female teams. Of those that reported work hours, 91.6 % reported working full time in condition A; 90% reported working fulltime in condition B.

Instructional activity design and implementation

As part of the course, students were required to read a chapter from the required text or supplementary materials each week. The pervasive practice for holding students accountable for readings in other sections of this course is requiring students to take multiple choice quizzes, but we wanted to provide opportunities for collaborative discourse. For this reason, students were assigned to teams in weeks three through five and in weeks six, eight, ten, twelve, and fifteen, we replaced required multiple choice quizzes with graded, synchronous discussions about reading content. On weeks when students had to complete the collaborative discussion, students had to set a meeting time with their teammates, and individually complete a pre-discussion activity and review the discussion session materials before the meeting. In session one, teams received full credit for the discussion regardless of the discussion quality. After the first session, students were given initial assessments and told that the subsequent discussions would be graded based on discussion quality.

The pre-discussion activity

We created a pre-discussion activity to help students organize their thinking around the required readings. The activity consisted five questions: (1) what were the main learning goals of the chapter, (2) what was the most difficult concepts or parts of the reading, (3) what did you find most interesting, (4) what four questions could you ask yourself, the authors, or others regarding this chapter, and (5) were you able to fully meet the learning goals for this chapter. Students had to respond to these questions and submit their responses before their discussion session.

The discussion materials

The discussion materials, housed in the LMS throughout the course duration, included a group process rubric detailing how we would assess discussion quality. We also provided students with guides containing goals and strategies associated with different collaborative capacities including information synthesis and knowledge negotiation.

The discussion sessions

The discussion sessions were held in a professional collaborative workspace with chat and document sharing capabilities. Use of video during discussion sessions was not allowed for two reasons: (1) many of our students lack access to high-speed Internet and (2) use of video would likely degrade chat quality. Each team had an assigned space that they could log in to and maintain for the duration of the course. Students were required to export their chat files after completing the discussion sessions. The exported files had to be submitted to a drop box folder in the LMS.

There were three parts to the discussion session, each with a different allotted amount of time for completion. Students were advised to spend no more than 1.5 hours to complete the entire three-part session. In part one, the team had 60-minutes to discuss questions and issues raised by the pre-discussion activity. Each team was also required to create and submit an outline of their discussion. In part two, the team had 15 minutes to individually assess the quality of their discussion session using the provided group process rubric, without communicating with other members. We informed students that an expert rater would assess the quality of their discussions and that we would determine the accuracy of their scores based on the difference between their scores and the expert score. The instructions stated, "it is more important to be accurate than it is to say your team did well. It will not help your team at all to give yourselves high scores. It is better to be critical, as this will help your team improve." Students were required to submit their individual assessments to a drop box in the LMS before moving on to part three. In part three, the team had 15 minutes to discuss how each team member assessed the team, identify their strengths and weaknesses, and select strategies from the materials provided that they could use to improve their collaborative discussion processes.

Assessment of collaborative discourse quality

We designed an assessment for determining the quality of online discussions by adapting a video-based collaborative interaction analysis rubric developed by Borge & Carroll (2010). After each discussion session, individual students evaluated the quality information synthesis and knowledge negotiation by completing this assessment. A research assistant with two years of communication analysis training used the same assessment to evaluate each team's discussion using their submitted chat files. There are three categories of behavior within each of the two core capacities, with each category assessed on a five-item, ordinal scale. Twenty percent of the total data was double coded by the research assistant and another trained graduate student to determine inter-rater reliability of the instrument: $r = 0.86$; $p < 0.001$, Kappa = 0.64; $p < 0.001$. Once each item of a core capacity is rated, they are averaged to produce a single Collaborative Discussion Quality score, which is a continuous value between 0 and 5 that we use to track improvement over time in collaborative discussion processes in the analysis below.

The first core capacity, information synthesis, consists of three categories of behavior. (1) *Verbal participation* examines the amount of turns of speech contributed by each member relative to the team's total turns of speech. Each chat message on the chat file is taken as a turn of speech. A score of one means that one member contributed most turns of speech and at least one member barely contributed. A score of five means all members contributed equally to the conversation. (2) *Developing joint understanding* evaluates the extent to which teams make an effort to ensure that members fully understand the ideas presented by taking time to reword, rephrase, or ask for further clarification of shared information. A score of one means that the team showed no instances of rewording, summarizing, or confirming another member's idea or decision, or a possible team action. A score of five means two criteria were met: (i) at least two instances exist where a member reworded another member's idea to make sure he/she understood it or asked another member to explain an idea by elaborating further, and (ii) at least one example exists of synthesizing major decisions or multiple ideas of members. (3) *Joint idea building* focuses on the extent to which team members elaborate on another member's contribution in order to ensure that information introduced by any member is not ignored or accepted, without discussion. A score of one means there were no instances where members extended or clarified another member's shared information; members ignored others, posed different suggestions unrelated to the original idea, or simply accepted the idea and moved on. A score of five means there were two or more instances where one or more members added to another's idea by extending or clarifying over more than five turns and there were no instances where members ignored others or posed different suggestions unrelated to the original idea.

The next core capacity, known as knowledge negotiation, also consists of three categories of behavior. (1) *Contributing alternative ideas* evaluates the extent to which teams present and discuss alternative perspectives, claims, or suggestions. A score of one means there were no instances where a claim or suggestion was followed by another member prompting for a counter claim, pointing out a problem, or sharing an alternative viewpoint. A score of five means that there were at least two examples where a claim was followed by another member prompting for a counter claim, pointing out a problem with a claim, or sharing an alternative viewpoint, and a discussion lasting over five turns ensues as a result. (2) *Quality of claims* focuses on evaluating the extent to which teams provide logical, fact-based evidence and rationale. A score of one means that when members made claims they did not include any rationale, evidence, or weighing of options. A score of five means there were at least two examples where claims were supported by logical, evidence-based rationale or weighing of different options. (3) *Norms of evaluation* focuses on evaluating the extent to which teams adhere to social norms that promote the development of psychological safety, "a shared belief held by members of a team that the team is safe for interpersonal risk taking" (Edmondson, 1999). A score of one means that members repeatedly used extremely inappropriate or offensive language (i.e., blatant profanity, vulgarity, racist or sexist language, etc.), or examples exist where a member attacks another member's intelligence or character (e.g. "you don't know what you're talking about"), or made disrespectful comments about a member's ideas (e.g. "that is stupid"). A score of five means that responses were professional and respectful with at least one instance where a member acknowledged that an opinion or claim of another member is reasonable or justifiable before pointing out its flaws or presenting a counter argument. Also, no examples exist where a member attacked another's intelligence or character, made disrespectful comments about an idea, or used inappropriate or offensive language (i.e., racist, sexist, or sexual in content).

Findings

(RQ1) What are the most common types of problems that interfere with high quality collaborative discourse practices?

In the first session, students completed the discussion activity based on their initial collaborative predispositions. We used expert ratings from this first session to identify the most common problems faced by our online population of students.

With regard to information synthesis, we found our online students were good at extending the ideas of teammates and developing shared understanding with teammates; these were the two highest average scoring areas (see table 1 for population means and modes by process). Though our population scored high on information synthesis overall, they experienced problems associated with verbal participation. Most of our teams, eight out of the thirteen teams, had one person dominate the majority of the team's discussion.

We found that our population of online students had more problems with collective knowledge negotiation than they did with collective information synthesis. The only item in knowledge negotiation that our population was able to perform at an average range was developing norms of constructive evaluation. Eleven out of the thirteen teams scored at an average level on this item, meaning that the teams primarily focused on evaluating ideas and not the people who suggested them. Teams were also not rude or hostile when evaluating claims. However, only two teams took time to acknowledge that an opinion or claim was reasonable or justifiable before pointing out flaws or presenting counter arguments.

When we examined the extent to which team members provided alternative ideas or counterclaims, we found that over half of our teams displayed processes at or below average. Three teams had slightly dysfunctional processes with regard to providing alternative ideas: people were prompting for counter claims or alternative viewpoints, but counter claims and alternative viewpoints were rejected or ignored without discussion. Four teams scored in the average range, meaning there was at least one example of someone prompting for or providing an alternative claim or opposing viewpoint and team members did not reject or ignore counter arguments or alternative viewpoints, but alternative claims or opposing viewpoint were immediately followed by agreement rather than discussion.

Table 1: The mean, mode, and standard deviation of each item of the assessment for our population. Teams were rated on a scale from one to five, where a score of two indicated some level of dysfunctional behavior

	Collective Information Synthesis			Collective Knowledge Negotiation		
	Verbal Participation	Joint Understanding	Idea Building	Alternative ideas	Quality of Claims	Norms of Evaluation
Mean	2.62	4.54	4.54	3.38	3.31	3.31
Mode	**2.00**	**5.00**	**5.00**	**2.00**	**2.00**	3.00
Std. Dev.	1.30	0.51	0.80	1.24	1.22	0.78

Quality of claims was also a problem for our teams. The most common score on the assessment was a score of two, indicating below average quality of claims. Teams with slightly dysfunctional quality of claims did not support any of the claims made during their session with logic-based rationale. These teams provided only shallow rationale; they did not weigh ideas or provide a chain of logical reasons to argue for or against a claim:

Turn	Speaker	Contribution
1	Tom:	Without databases all the logistics of our planet today would be a nightmare.
2	Pete:	Yes it makes everything way more efficient
3	Rob:	Yeah I can't imagine a world without databases
4	Tom:	I think the chapter could have touched on more SQL language bits
		it was mentioned for about 5 seconds and then gone

The team makes four claims in each turn of this short episode. Turns 1-3 include a claim with no rationale or evidence. In Turn 4, Tom makes the claim that the chapter could have discussed SQL language more in depth, but supports this claim with a shallow, opinion-based rationale.

An average score on quality of claims would denote a pattern of claim making where the team supports their claims with logical, but opinion-based rationale. To score above average (score of four), teams had to show evidence of at least one instance of claims supported with logical, evidence-based rationale that referred to course content from the text or another information source such as the one below:

> "I beg to differ just a little bit, Hal. It's not so much a lack of security as it is the fact that absolutely anything is hackable. Remember how the first chapter of the book discussed bits, everything is 1's and 0's to a computer? You can reverse engineer computer code, break it down to those most basic of components, so there is really no such thing as perfect security. The only sure-fire way to keep your information safe is to never share it to begin with."

Only two out of our thirteen teams scored top marks on this item, having two or more examples of logical, evidence-based rationale during their one-hour discussion session.

(RQ2) To what extent does the activity design model and differing reflective activities affect teams' improvement of collaborative discourse over time?

In refining our model, we wanted to (1) test the utility of our activity design model as a means to support the improvement of the quality of collaborative discussion processes over time and (2) determine which of two reflective procedures best facilitated sociometacognitive learning: students' ability to modify activity to meet collaborative process goals. To accomplish these aims, teams were placed into one of two individual reflective assessment conditions and a trained research assistant measured Collaborative Discussion Quality at five time points. All teams followed the activity design model, but there were two conditions for reflective activity. After each discussion session, teams assigned to condition A, future thinking, were required to score their team using the Collaborative Discussion Quality assessment and then provide a strategy they could use in the next session to improve the quality of their collaborative processes. Teams assigned to condition B, evidence, followed the same procedures as condition A with one exception: they provided evidence from the discussion to support their self-assessment ratings instead of a providing strategy to improve performance.

Altogether, the data set consisted of 125 data points. We built an ANOVA model with team Collaborative Discussion Quality as measured by an expert rater at a time point as the dependent measure. Condition was the independent variable. Time nested within Condition was a covariate. Additionally, we included Team nested within Condition as a control variable. If the manipulation supports quality at each time point, we expect a main effect of condition. If quality improves significantly over time, we expect to see an effect of time. If the improvement is different between the two conditions, we expect the slope associated with the time variable to differ between conditions.

We found Time had a significant effect in the model. The partial correlation of Time on Discussion Quality in this model was .45, $p < .005$. This suggests that the activity design model facilitated the improvement collaborative discussion quality over time.

The effect of Condition was also significant: $F(1, 110) = 5.46$, $p < .05$, effect size .37 standard deviations. Teams in Condition A (future thinking) had lower scores on average ($M = 11.07$, $SD = 2.19$) than teams in Condition B (evidence) ($M = 11.87$, $SD = 2.11$). The slope for Condition A was slightly higher than for Condition B, but the difference in slopes was not statistically significant when compared through a hierarchical growth model. In this analysis, both Condition and Time were significant predictors of collaborative process quality. Though teams improved on both collective information synthesis and knowledge negotiation, we found that most of the improvement was due to students' consistent improvement with the quality of knowledge negotiation. In session one, ten teams were rated as below average on knowledge negotiation processes, but the number of teams scoring below average steadily decreased over time. By session four, four teams were below average. No teams were below average in session five. The same pattern was not true of information synthesis.

Conclusions and implications

The work presented in this paper helps to inform the design of activities that could be included as part of CSCL systems to support collective regulation and improvement of collaborative processes. Our online students were prone to similar types of dysfunctional patterns of collaborative interaction as undergraduate students in face-to-face instructional conditions (Borge & Carroll, 2010; Carroll et al, 2014). Initially, online teams were able to discuss course concepts, share opinions, and extend the ideas of others, but the diversity and quality of the claims that were made, along with verbal participation, were less than optimal.

Rather that guide students' collaborative activity during discussion sessions, we chose to provide students with reflective assessment activities. These activities articulated a model of optimal collaborative processes for collective information synthesis and knowledge negotiation. We found that this approach combined with feedback succeeded in helping students to improve the quality of their discussions over time, thus supporting the utility of the activity design model. However, we have yet to fully investigate students' perspectives of the utility of the activity and analyze their feedback to inform the model.

With regard to evaluating the two different types of reflective activities we found that asking students to provide evidence from their chat sessions to support their assessments of process quality (evidence condition) was associated with more improvement over time than asking students to provide strategies they could use to improve future performance (future thinking condition). This is interesting because we expected that requiring students to propose strategies for future discussions would facilitate behavior change and lead to more improvement, but our data does not support this claim. One explanation for the difference between reflective assessment conditions is that students may need more support when thinking about current processes, before moving onto future planning. It is possible that requiring students to provide evidence for their ratings from their discussion sessions pushed them to think more critically about how they were assessing their team and whether they understood the assessment items. In thinking more deeply about the assessment, students may internalize the assessment criteria more than they would otherwise.

We are working to incorporate our findings to inform the design of a CSCL environment with awareness affordances designed to support reflection as online students work to collectively regulate collaborative interactions. Future studies will explore how to help students improve their accuracy of assessment over time. Preliminary experimentation with automated analysis of the discussion data suggests that while machine learning models for assessment are less accurate than experts, they may be more accurate than the students, and thus it may be feasible to use this technology to support improvement in this area. Investigating how to improve this automated assessment and use it in interventions for supporting improvement in self-assessment accuracy over time is an important direction of our continued research.

References

Barron, B. (2003). When smart groups fail. Journal of the Learning Sciences, 12(3), 307-359.

Bjork, R.A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), Metacognition: Knowing about knowing (pp. 185–205). Cambridge, MA: MIT Press.

Boekarts, M. (1996). Coping with stress in childhood and adolescence. In M. Zeidner & N.S. Endler (Eds.), Handbook of coping: Theory, research, and applications (pp. 452–484). New York: Wiley.

Borge, M. & Carroll, J.M. (2010). Using collaborative activity as a means to explore student performance and understanding. In Gomez, K., Lyons, L., & Radinsky, J. (Eds.) Learning in the Disciplines: Proceedings of ICLS 2010: 9th International Conference of the Learning Sciences (Chicago, June 29-July 2) Volume 1, Full Papers. International Society of the Learning Sciences: Chicago IL.

Borge, M., Ganoe, C. H., Shih, S. I., & Carroll, J. M. (2012). Patterns of team processes and breakdowns in information analysis tasks. In Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (pp. 1105-1114). ACM.

Burson, K. A., Larrick, R. P., & Klayman, J. (2006). Skilled or unskilled, but still unaware of it: how perceptions of difficulty drive miscalibration in relative comparisons. Journal of personality and social psychology, 90(1), 60.

Carroll, J. M., Jiang, H., & Borge, M. (2014). Distributed collaborative homework activities in a problem-based usability engineering course. Education and Information Technologies, 1-29.

DeShon, R. P., Kozlowski, S. W., Schmidt, A. M., Milner, K. R., & Wiechmann, D. (2004). A multiple-goal, multilevel model of feedback effects on the regulation of individual and team performance. Journal of applied psychology, 89(6), 1035-1056.

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

Edmondson, A. C. 1999. Psychological safety and learning behavior in work teams. Administrative Science Quarterly, 44(2) 350-383.

Fischer, F., Kollar, I., Stegmann, K., Wecker, C., Zottman, J., & Weinberger, A. (2013). Collaboration Scripts in Computer Supported Collaborative Learning, in Hmelo-Silver, C., Chinn, C., Chan. C., & O'Donnell, A. (Eds) The International Handbook of Collaborative Learning. New York, NY: Routledge.

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

Hogan, K. (1999b). Thinking aloud together: A test of an intervention to foster students' collaborative scientific reasoning. Journal of Research in Science Teaching, 36(10), 1085-1109.

Järvelä, S., & Hadwin, A. F. (2013). New frontiers: Regulating learning in CSCL. Educational Psychologist, 48(1), 25-39.

Kapur, M., & Rummel, N. (2009). The assistance dilemma in CSCL. In Proceedings of the 9th international conference on Computer supported collaborative learning-Volume 2 (pp. 37-39). International Society of the Learning Sciences.

Kerr, N. L., & Tindale, R. S. (2004). Group performance and decision making. Annual Review of Psychology, 55, 623-655.

Kozlowski, S., & Ilgen, D. (2006). Enhancing the effectiveness of work groups and teams. Psychological Science in the Public Interest, 7(3), 77-124.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. Journal of personality and social psychology, 77(6), 1121.

Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. Science, 328(5977), 463-466.

Schoenfeld, A. H. (1987). What's all the fuss about metacognition. In A. H. Schoenfeld (Ed.), Cognitive science and mathematics education (pp. 189-215). Hillsdale, NJ: Lawrence Erlbaum Associates.

Schraw, G. (1998). Promoting general metacognitive awareness. Instructional science, 26(1-2), 113-125.

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

Shimoda, T., White, B., Borge, M., & Frederiksen, J. (2013). Designing for science learning and collaborative discourse. In Proceedings of the 12th International Conference on Interaction Design and Children (pp. 247-256). ACM.

Stegmann, K., Mu, J., Gehlen-Baum, V., & Fischer, F. (2011). The myth of over-scripting: Can novices be supported too much. In Connecting Computer-Supported Collaborative Learning to Policy and Practice: CSCL2011 Conference Proceedings (Vol. 1, pp. 406-413).

Webb, N. M., & Palincsar, A. S. (1996). Group processes in the classroom. Englewood Cliffs NJ: Prentice Hall International.

White, B., & Frederiksen, J. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. Cognition and Instruction, 16(1), 3-118.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), Metacognition in educational theory and practice (pp. 277e304). Mahwah, NJ: Erlbaum.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. Theory into practice, 41(2), 64-70. http://psycnet.apa.org/doi/10.1037/10096-015

Acknowledgments

We would like to thank Todd Shimoda for his valuable input and our undergraduate research assistants, Emily Hanson and Scott Cunningham, for their contributions to this project. We would also like to thank the participating students for allowing us to examine their interactions and for giving us constructive, thoughtful feedback on the activities. This project was supported by the National Science Foundation (IIS-1319445).