Maximizing learning from collaborative activities

Rachel J. Lam (rachel.lam@asu.edu)
Arizona State University
Teachers College, 300 E. Lemon St.
Tempe, AZ 85281

Abstract

Utilizing a Preparation for Future Learning paradigm and the Interactive-Constructive-Active-Passive framework, this study examined how two different kinds of cognitively engaging activities prepared students to learn from collaborating. Findings show that preparing prior to collaborating improved learning, but a difference was not detected in the type of preparation. In addition, differences in learning outcomes were only present in measures of deep knowledge. Analyses used a multilevel method targeted to dyadic data. Discussion addresses designing collaborative classroom activities that are effective and efficient for deep learning, as well as the importance of aligning assessments to depth of learning.

Keywords: collaborative learning; preparation for future learning; cognitive engagement; classroom learning.

Introduction

Collaborative learning has become a common instructional strategy in a variety of educational settings because of its potential to boost student learning. Through peer discussion, students can receive immediate feedback, ask questions, generate explanations, challenge each other, jointly construct understanding, and elaborate on each other's ideas, which are all behaviors that have been shown to improve learning outcomes in both the classroom and laboratory. However, despite the extensive research that has been conducted on collaborative learning, the literature is still unclear as to what factors lead to the best learning outcomes, in particular, for deep understanding of concepts. Thus, this work aimed to investigate two factors that may improve deep knowledge, in particular, in a conceptual (as opposed to a problem-solving) domain: (a) individually engaging in the learning material prior to collaborating and (b) "constructively" engaging, where students are generating (constructing) new knowledge beyond the learning material.

There are mixed results as to how collaboration affects student learning (Barron, 2003; Craig, Chi, & VanLehn, 2009). In general, students do not always take advantage of the benefits collaboration affords, thus, researchers have searched for ways to help students collaborate more effectively. Methods such as training students in collaboration skills (Hausmann, 2006; Uesaka & Manalo, 2011), providing structured guidance to students while interacting (Coleman, 1998; Walker, Rummel, & Koedinger, 2011), and designing collaborative learning environments that elicit meaningful discussion (Engle & Conant, 2002; Kapur & Bielaczyc, 2012) have been found to improve learning from collaborating. However, there are also challenges and limitations to these methods.

One limitation to training students in specific skills before collaborating is that they often fail to retain those skills after time (Webb, Nemer, & Ing, 2006). The challenge of structured guidance during collaboration is that too much can constrain creativity and flexible discussion, which can hinder learning (Cohen, 1994). Therefore, one question that remains is, does the effort and time that it takes to train or guide students in collaborative behaviors really pay off? Work that has investigated the design of collaborative activities to naturally elicit effective dialoguing addresses this challenge, showing that open-ended and flexible tasks can enrich discussion (Janssen, Erkens, Kirshner, & Kanselaar, 2010; Van Boxtel, Van der Linden, & Kanselaar, 2000). However, this only occurs when students have sufficient prior knowledge (Nokes-Malach, Meade, & Morrow, 2012). Thus, a collaborative learning method that avoids the time and effort needed to train students in particular skills or structure their instance-by-instance dialogic behaviors, while providing the opportunities for students to acquire adequate prior knowledge is investigated in the current study.

Cognitive theoretical models

Two cognitive theoretical models supported the design of the collaborative activities in this study. The Interactive-Constructive-Active-Passive (ICAP) framework and the Preparation for Future Learning (PFL) paradigm are described below.

The ICAP framework

The ICAP framework differentiates student engagement in learning tasks by categorizing students' overt behaviors as Interactive, Constructive, Active, or Passive, and is founded on theoretical assumptions about how those behaviors link to different cognitive processes (Chi, 2009; Menekse, Stump, Krause, & Chi, 2012). An Interactive behavior might be debating or extending a partner's idea and the cognitive process underlying Interactive engagement would be co-creating knowledge. Inventing a rule, self-explaining, or creating a concept map would be Constructive, the underlying cognitive process being creating new knowledge. Active behaviors include highlighting a textbook chapter or copying solutions steps from the board, and correspond to assimilating knowledge. Listening or watching would be considered Passive, corresponding to the process of storing knowledge. The ICAP hypothesis makes the prediction that Interactive activities will produce better learning outcomes than Constructive activities, which are better than Active

activities, which are all better than Passive activities: I>C>A>P. There is empirical support for the ICAP hypothesis, although the Interactive category carries several caveats (Menekse et al., 2012). One is that engagement should only be considered Interactive when both individuals in a dialogue are engaging constructively. This does not always occur (literature on the process of collaboration in learning settings attests to this claim). Thus, this current study will address the question of how learning is affected by interacting on a Constructively designed task or an Actively designed task.

The PFL paradigm

This paradigm takes into account how earlier learning experiences can shape future learning, under the perspective that prior learning can activate a mental model to either facilitate or hinder the learning of a new concept (Schwartz, Sears, & Chang, 2007). Although the PFL paradigm was introduced in the literature over two decades ago (Schwartz & Bransford, 1998), more recent work has used this model to investigate learning outcomes in a variety of domains (Chin et al., 2010, in elementary school science; Gadgil & Nokes-Malach, 2012, in cognitive psychology; Schwartz, Chase, Oppezzo, & Chin, in press, in physics). This work has shown that invention-type tasks better prepare students to learn from a lecture (Schwartz & Martin, 2004). In other words, tasks that are set up to cognitively engage students in a "constructive" way, by causing students to generate new knowledge (Chi, 2009), are those that best prepare students to learn in a future task. The majority of the work that has investigated the PFL paradigm uses some form of didactic instruction (i.e. lecture) as the future task, thus, little is known about the effects other forms of instruction as future tasks, such as collaboration. The current study utilizes the PFL model to structure collaborative learning activities for students, however, the future activity is peer discussion (instead of a lecture) and students individually (rather than collaboratively) engage in the preparation task.

Measures of learning and mental models

In light of using the two aforementioned cognitive perspectives as the basis for this study, the measures of learning outcomes should be viewed as representing student mental models of the concepts being tested. Mental models can be assessed through externalizations such as self-generated concept maps, matrices, drawings, and free-writing (Janssen et al., 2010; Schwartz, 1995; Van Amelsvoort, Andriessen, & Kanselaar, 2007). Multiple-choice or T/F tests are often used to measure student learning with regard to accuracy or correctness of knowledge, however, these are not necessarily appropriate to fully assess a mental model (Bransford & Schwartz, 1999; Schwartz et al., 2007). A more complete picture of student knowledge can be captured by combining these types of assessments. With respect to measuring depth of knowledge, shallow knowledge can be equated to the "surface features" of a mental model, while deep knowledge

lies in the "structure" of the model (Chi & VanLehn, 2012). Surface features can be facets such as labels and definitions, physical characteristics, or other plain facts. Structural knowledge is much more complex, representing the relationships between the features of a concept and/or the process by which a concept occurs or functions. Thus, the current work used student-generated written responses to assess deep, structural-based learning, while T/F pre- and posttests were used to assess shallow, surface-feature learning.

Method

The study used a 2x2 experimental design examining Preparation (No Prep and Prep) and Type of Task (Active and Constructive). The two dependent variables were shallow learning and deep learning. In order to preserve both internal validity and ecological validity, the study was conducted as a classroom study across four introductory psychology classes with equal representation of the four conditions in each classroom. The students participated in the study as a part of their "regular" classroom activity for the weekly topic of "concepts of memory."

Participants

Ninety students from four Psych 101 courses at a large community college in a Southwestern city in the United States participated in this study. The mean age of students was 21 years and the sample represented an ethnically diverse population (46% Hispanic, 37% Caucasian, 10% African American, and 7% Asian, Native American, or Middle Eastern). Fifty six percent of the students were female, 44% were male.

Materials

Regarding the topic of interest, prior research attests to the difficulty that students have in deeply understanding the differences between a variety of concepts of memory, in particular, for encoding- and schema-based concepts (Schwartz & Bransford, 1998). Thus, all learning activity materials and assessments were based on Schwartz and Bransford's (1998) materials. These materials were the only form of instruction to students for the topic. Students received no other instructional material (lecture, textbook readings, etc.) prior to the study and, therefore, were assumed to have limited prior knowledge of the concepts.

The study used the following materials: (1) pretest and demographic survey, (2) four versions of learning materials based on condition, (3) posttest, and (4) scoring rubrics.

(1) The pretest consisted of T/F questions that were very slightly modified from Schwartz and Bransford's (1998) verification measure, which was used in several studies on concepts of memory.

(2) The materials used during the learning phase were equivalent in domain content, however, the specific task instructions varied according to the ICAP cognitive engagement definitions and whether or not the condition

included a preparation period. In Prep conditions, students were given a portion of the class time to individually work on the task prior to engaging with a partner, while students in the No Prep conditions worked with a partner for the entirety of the learning phase. Active tasks asked students to work within the existing learning materials (i.e. they did not have to generate inferences beyond the materials to complete the tasks), while the Constructive tasks required students to invent concepts. To provide an example, the Constructive task required students to answer questions such as, "Why do people remember certain kinds of information, but not other kinds?" after studying a memory experiment and its results. They had to generate ideas about the process of memory. The Active version of the task, on the other hand, instructed students to study a list of memory terms and their descriptions. They then applied the terms to the same memory experiment included in the Constructive version by writing the term next to the appropriate result of the experiment. These students had to "search and select," but did not necessarily have to generate any new knowledge. Since the Active tasks took much less time to complete (as shown in a prior pilot study of this work), they included a secondary memory experiment task that was identical in structure to the first, but with a different cover story. This was to control for time-on-task, which was equalized across the four conditions.

(3) The posttest included the same T/F questions that were used in the pretest. To avoid a "testing effect" (i.e. learning solely attributed to the recognition of identical test questions at a later testing phase), the ordering of the questions was changed and there were four to five days in-between the tests. (See work by Bjork and Storm, 2011, for details regarding the conditions under which testing influences learning.) Student gain scores from pre- to posttest served as the measure of shallow learning.

Two additional tasks were included on the posttest to obtain a measure of deep learning. These were "prediction" tasks, where students had to study novel experiments on memory (i.e. they did not appear in the learning materials) and synthesize their recently learned knowledge in order to apply it to new experimental conditions, generate new inferences about how memory works, predict the results of the experiments, and provide evidence of their reasoning for predictions. Students freely wrote their responses to a set of sub-questions that all corresponded to a basic question of, "Based on what you now know about memory, how do you think the results of these experiments will turn out?"

Because these types of prediction tasks are likely deeply cognitively engaging, there was concern that including any on the pretest might influence students to engage differently in the learning activity tasks. In particular, the Active conditions may have become contaminated if students were primed in a pretest task to think more deeply about the concepts. Thus, the pretest only included the shallow T/F questions. Although this prevented obtaining any measure of deep knowledge prior to the learning phase, this was of less concern since it was highly unlikely that students had

prior deep knowledge of memory concepts. As already mentioned, they not did have previous instruction on the topic in their classes and in addition, they produced low shallow knowledge scores at pretest (M=50.8%, SD=21.6). Thus, rather than a gain score, the deep learning measure used only the posttest prediction task scores.

(4) Scoring rubrics were developed in order to quantify students' responses to these prediction tasks. Responses were coded by how well they represented any of the following eight concepts: elaboration, schema, gist, serial position effect, generation effect, obstacle recall, interference, and encoding failure. These concepts may have been explicitly learned in the Active conditions, through the "search and select" tasks, or may have been implicitly learned in the Constructive conditions, through the "invention of concepts" tasks. A code of "other" was used for responses that represented novel ideas about memory (i.e. ideas that were not taught through the activities). This coding translated to a score ranging from 0-3 points, based on a holistic-style rubric. A higher score indicated knowledge of a broader range of concepts, representing a more complete mental model of memory. A score was also given for the quality of students' reasoning supporting the relationship between their predictions and the concepts, also ranging from 0-3 points. This score indicated knowledge of the relationships between the concepts and their applications to novel settings, thus, representing a better structured mental model. A total score of 0-6 was possible. Two raters scored a randomly selected 20% of the data and intraclass correlation was used to assess inter-rater reliability, ICC(2,1)=.76, p<.001. One rater scored the remaining tests.

Procedure

The study took place over the course of a week. On the first day, students took the pretest and filled out the demographic survey. Students were given 15 minutes to complete the pretest.

Students completed the learning activity phase during the next class. They were randomly assigned to one of the four conditions: (a) No Prep-Active, (b) No Prep-Constructive, (c) Prep-Active, and (d) Prep-Constructive. For No Prep conditions, students were randomly assigned to a partner and told to follow the instructions on their packets. They were encouraged to share ideas, try come to agreement before writing down an answer, and not to worry about writing right or wrong answers. Instructions varied depending on whether students were completing the Active or Constructive version (described in the Materials section), but all students were told to try to contribute equally to the discussion. For Prep conditions, students first completed an individual packet. They were told not to worry about right or wrong answers and to do their best. They were informed that they would use this packet to work with a partner. After the individual work (ranging from 15-20 minutes), students were randomly paired and spent the remaining class period doing their collaborative packet (10-15 minutes). They were told to share their ideas, try to contribute equally to

discussion, and come to agreement before writing down their answers. At the end of class, all materials were collected (each pair turned in a jointly completed collaborative packet). Students spent 30-35 minutes on the learning task in all conditions.

The posttest was given in the following class and was completed individually. Students spent 35-50 minutes on the posttest. Any students who finished before 30 minutes passed were asked to go over their answers one more time.

Results

To avoid violation of the assumption of independence of subjects (which traditional ANOVA assumes) (Kenny, Kashy, & Cook, 2006), a dyadic multilevel model was used for all analyses. Figure 1 illustrates the model. The analytic technique was a linear mixed model with the Restricted Maximum Likelihood (REML) method, appropriate to cope with dependency between partners within dyads.
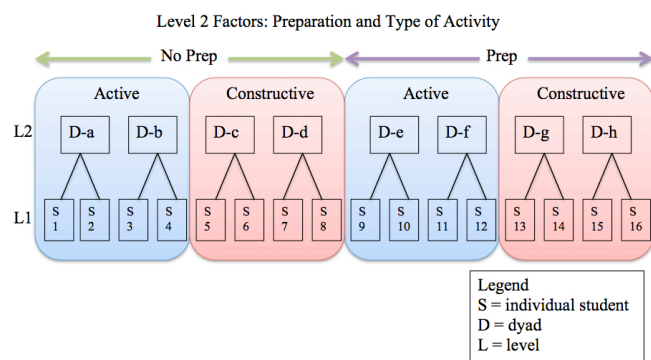

Figure 1: Multilevel dyadic design.

Shallow learning

Analysis of the pre- and posttests compared learning gains across conditions. "Normalized change" calculations were used to adjust learning gains by accounting for influences of pretest scores, yielding a more sensitive measure of gains (Marx & Cummings, 2007). When post>pre, the following formula was used: post-pre/1-pre. When post<pre, a different formula was used: post-pre/pre. Although students gained in all conditions, there was no reliable difference between conditions. Table 1 summarizes these results.

Table 1: Shallow learning mean scores

Condition	n	Pretest%	Posttest%	Adj. Gain
No Prep-Active	14	53.6	72.6	.43
No Prep-Constructive	18	49.1	61.1	.21
Prep-Active	15	46.7	63.3	.28
Prep-Constructive	19	53.5	71.9	.40
Total	66	50.8	67.2	.33

Note: Due to incompletion of the T/F questions at either pre- or posttest, the total sample was reduced from 90 to 66 students.

These results are not surprising because even the "lowest" condition (No Prep-Active) constitutes an effective teaching strategy in a number of ways. Students were provided terms and definitions, the opportunity to apply those to real-world examples, and the benefit of engaging in discussion. Because these pre- and posttests were used to assess the knowledge of the surface features of memory concepts, students were expected to gain in all conditions. The differences between conditions were only hypothesized for deep learning, attesting to the sensitivity of the manipulation of the conditions. The deep learning results are below.

Deep learning

Ninety students completed the prediction task portion of the posttest. The prediction task posttest scores were reliably different across conditions. There was a main effect of Preparation $F(1,41.1)=5.79$, $p<.03$, but no effect of Type of Task, nor an interaction effect. Students who prepared in the task individually in either type of task before collaborating showed evidence of deeper learning. See Figure 2.
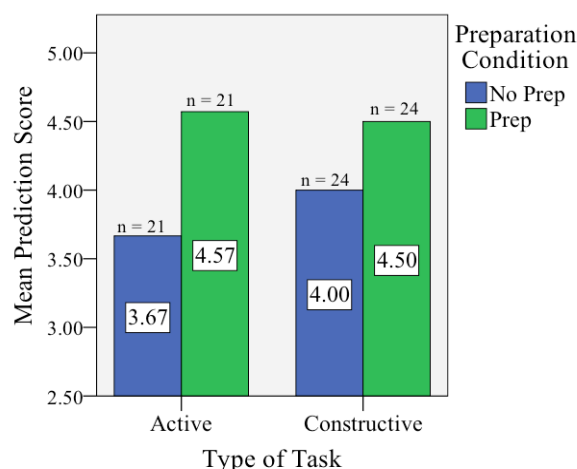

Figure 2: Prediction task results.

This result was not expected since prior work supports the notion that "constructively" engaging activities should produce improved learning above "actively" engaging activities. As shown in Figure 2, there is virtually no difference between the Active and Constructive conditions when students individually prepared prior to collaborating. One interpretation of these results is that the inclusion of preparation prior to discussion in a collaborative activity boosts learning such that it overrides any effects of type of task. It is possible that the inclusion of an individual preparation period increases the likelihood that students will engage constructively in a dialogue, regardless of whether the task itself requires generation of new knowledge. In other words, the preparation may have spontaneously impelled students to engage constructively even in Active tasks, thus, further exploratory analyses were conducted to check this.

Preparation facilitates constructive engagement

To assess differences between how the No Prep-Active and Prep-Active students engaged in the tasks, the collaborative activity worksheets were examined. They were scored by student effort, rather than in correctness of responses, since these were never intended to measure learning. Support for such a strategy can be found in work on dynamic assessments (Bransford & Schwartz, 1999; Schwartz et al., 2007), which measure readiness to learn, rather than learning outcomes. Thus, as related to the PFL paradigm, these learning tasks can be viewed as readiness tasks that prepared students to engage in collaboration, with the posttest prediction task measuring learning. It is possible that students from the Prep-Active condition developed an enhanced readiness for learning in discussion, accounting for the improved performance on the prediction tasks.

Each dyad that completed at least 94% (15 of 16 items) of the activity worksheet received an effort score of two; those that completed at least 75% received an effort score of one; zero points were given to any dyads that completed under 69% of the activity (only four dyads). Since amount of work completed is not a thorough indication of how deeply engaged students were in the activities, the number of times students within a pair disagreed was also taken into account. The worksheets included a line for each item that asked students if they agreed on the answer, and if not, they were instructed to explain their disagreements. (Work on argumentation shows that students benefit from talking through disagreements, Asterhan & Schwarz, 2009.) Analysis of discourse could have provided a better measure of engagement, however, that was beyond the scope of this paper. Thus, activity effort and average number of disagreements per pair were used to measure engagement.

Results showed that dyads in the Prep-Active condition produced a higher activity score ($M = 1.71$) compared to those in the No Prep-Active condition ($M = 1.43$), and had a slightly higher average number of disagreements (.55 compared to .45, respectively). Although none of these differences were significant, put together they provide some support that preparation may influence students to engage constructively in an activity, even when the activity in and of itself does not require such engagement.

Discussion and future work

This study tested the effects of preparation and type of task on shallow and deep learning in a collaborative activity. Students engaged in either an Actively or Constructively designed task, and either worked individually during part of the learning phase, then collaborated (Prep), or worked jointly the entire time (No Prep). (Recall that time-on-task and domain content of the learning materials were the same across all conditions.) Results showed that preparation improved deep learning outcomes, but no difference was detected for type of task. The main effect of preparation on outcomes extends the PFL paradigm, showing that peer discussion can serve as a beneficial future learning task (i.e. the future task need not be lecture). Considering the learning opportunities that peer discussion offers as compared to didactic forms of instruction, this is an important finding towards design of classroom activities, especially with regard to deep learning. Although this study cannot inform on the comparison between collaborative learning and didactic instruction as future learning tasks of a PFL model, it supports the need for more work in this area.

The ICAP framework was not necessarily supported as an effective tool for designing learning activities since, overall, there were no differences in type of task on learning. However, the ICAP hypothesis predicts outcomes based on student engagement, not on task instructions. Thus, the exploratory analysis showed that students in the Prep-Active condition might have engaged constructively, justifying a null effect. Additionally, one might argue that because all four conditions included collaborative activities, the level of engagement for all conditions was actually Interactive. What is of interest here is that there then should have been an overall null result, however, that did not occur in the deep learning outcomes. Chi (2009) discusses the idea that working in pairs does not automatically make engagement Interactive, and that to be truly Interactive, both students must at minimum be engaging constructively. Thus, with regard to design of learning tasks, one way to better ensure that students engage Interactively in collaborative tasks is to include an individual preparation task prior to discussion. Future work is examining discourse data from a sampling of pairs from this study to further inform on how discourse processes related to learning, within in the contexts of the ICAP framework and PFL paradigm.

This study draws concern toward prior work that has not used analytic techniques that account for dependency between partners within dyads. Future work in areas of collaborative learning should utilize dyadic or multilevel models to analyze data that includes individual student outcomes (such as individually completed posttests).

Regarding the learning assessments, this study shows the usefulness of distinguishing between deep and shallow learning, and that different kinds of measures are needed to evaluate learning of varying depths. By using a mental model perspective to understand outcomes, one can see that a measure of "surface feature" knowledge would have shown no effects across conditions. An appropriate measure of deep "structural" knowledge was needed to tease apart how learning was affected by the collaborative tasks.

To conclude, it appears that one way to maximize the benefits of collaboration on deep learning is to include a preparation task, which allows students to develop a readiness for learning in future discussion. Preparation also may elicit spontaneous constructive engagement in future discussion, and it may not be necessary to otherwise design collaborative activities to specifically engage constructive behaviors. In addition, students who prepared only spent half the amount of the time collaborating. Thus, using a PFL paradigm to structure collaborative activity is also efficient, in that students can make the most effective use of their time engaging in discussion.

References

Asterhan, C.S.C. & Schwarz, B.B. (2009). Argumentation and explanation in conceptual change: Indications from protocol analyses of peer-to-peer dialog. *Cognitive Science, 33*, 374-400.

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

Bjork, E.L. & Storm, B.C. (2011). Retrieval experience as a modifier of future encoding: Another test effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37(5)*, 1113-1124.

Bransford, J.D. & Schwartz, D.L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education, 24*, 61-100.

Chi, M.T.H. (2009). Active-Constructive-Interactive: A conceptual framework of differentiating learning activities. *Topics in Cognitive Science, 1(1)*, 73-105.

Chi, M.T.H. & VanLehn, K. (2012). Seeing deep structure from the interactions of surface features. *Educational Psychologist*, *47(3)*, 177-188.

Chin, D.B., Dohmen, I.M., Cheng, B.H., Oppezzo, M.A., Chase, C.C., & Schwartz, D.L. (2010). Preparing students for future learning with Teachable Agents. *Educational Technology Research and Development, 58*, 649-669.

Cohen, E.G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research, 64(1)*, 1-35.

Coleman, E.B. (1998). Using explanatory knowledge during collaborative problem solving in science. *Journal of the Learning Sciences, 7(3&4)*, 387-427.

Craig, S.D., Chi, M.T.H. and VanLehn, K. (2009). Improving classroom learning by collaboratively observing human tutoring videos while problem solving. *Journal of Educational Psychology, 101(4)*, 779-789.

Engle, R.A. & Conant, F.R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. *Cognition and Instruction, 20(4)*, 399-483.

Gadgil, S. & Nokes-Malach, T.J. (2012). Overcoming collaborative inhibition through error correction: A classroom experiment. *Applied Cognitive Psychology*, DOI: 10.1002/acp.1843

Hausmann, R.G.M. (2006). Why do elaborative dialogs lead to effective problem solving and deep learning? *Proceedings of the 28th Annual Meeting of the Cognitive Science Society* (pp. 1465-1469). Vancouver, B.C.: Sheridan Printing.

Janssen, J., Erkens, G., Kirschner, P.A., & Kanselaar, G. (2010). Effects of representational guidance during computer-supported collaborative learning. *Instructional Science, 38(1)*, 59-88.

Kapur, M. & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences, 21(1)*, 45-83.

Kenny, D.A., Kashy, D.A, & Cook, W.L. (2006). *Analysis of Dyadic Data*. New York: Guilford.

Marx, J.D. & Cummings, K. (2007). Normalized change. *American Journal of Physics, 75(1)*, 87-91.

Menekse, M., Stump, G.S., Krause, S., & Chi, M.T.H. (in press). Differentiated overt learning activities for effective instruction in an engineering classroom. *Journal of Engineering Education*.

Nokes-Malach, T.J., Meade, M.L., & Morrow, D.G. (2012). The effect of expertise on collaborative problem solving. *Thinking & Reasoning, 18(1)*, 32-58.

Schwartz, D.L. (1995). The emergence of abstract representations in dyad problem solving. *The Journal of the Learning Sciences, 4(3),* 321- 354.

Schwartz, D.L. & Bransford, J.D. (1998). A time for telling. *Cognition and Instruction, 16(4)*, 475-522.

Schwartz, D.L., Chase, C.C., Oppezzo, M.A., & Chin, D.B. (in press). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology*.

Schwartz, D.L. & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction, 22(2)*, 129-184.

Schwartz, D.L., Sears, D., & Chang, J. (2007). Reconsidering prior knowledge. *Carnegie Symposium on Cognition*, 319-344. Mahwah, NJ: Lawrence Erlbaum.

Uesaka, Y. & Manalo, E. (2011). The effects of peer communication with diagrams on students' math word problem solving processes and outcomes. *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society* (pp. 312-317). Austin, TX: Cognitive Science Society.

Van Amelsvoort, M., Andriessen, J., & Kanselaar, G. (2007). Representational tools in computer-supported collaborative argumentation-based learning: How dyads work with constructed and inspected argumentative diagrams. *Journal of the Learning Sciences, 16(4)*, 485-521.

Van Boxtel, C., Van der Linden, J., & Kanselaar, G. (2000). Collaborative learning tasks and the elaboration of conceptual knowledge. *Learning and Instruction, 10*, 311-330.

Walker, E., Rummel, N., & Koedinger, K.R., (2011). Designing automated adaptive support to improve student helping behaviors in a peer tutoring activity. *Computer-Supported Collaborative Learning, 6*, 279-306.

Webb, N.M., Nemer, K.M., & Ing, M. (2006). Small-group reflections: Parallels between teacher discourse and student behavior in peer-directed groups. *Journal of the Learning Sciences, 15(1)*, 63-119.