Diversity, Collaboration, and Learning by Invention

Jennifer Wiley (jwiley@uic.edu)
Department of Psychology, 1007 W. Harrison Street
Chicago, IL 60607 USA

Olga Goldenberg
Department of Psychology, 1007 W. Harrison Street
Chicago, IL 60607 USA

Andrew F. Jarosz
Department of Psychology, 1007 W. Harrison Street
Chicago, IL 60607 USA

Michael Wiedmann
Institut für Psychologie, Engelbergerstr. 41
D-79085 Freiburg DE

Nikol Rummel
Institut für Erziehungswissenschaft, Universitätsstraße 150
D-44801 Bochum DE

Abstract

Learning-by-invention is an approach to mathematical instruction where small groups explore possible methods of solution before learning the "right answer" (e.g., Schwartz & Martin, 2004; Kapur & Bielaczyc, 2011). In a series of studies we have been investigating the effects of group composition in terms of math ability on learning by invention. An initial result showed that groups consisting of a mix of both high and low math ability students generated a broader range of solution attempts when asked to invent a formula for standard deviation compared to more homogeneous math ability groups. Moreover, this wider range of solution alternatives predicted better performance on quizzes following a lesson on the topic. Subsequent work is suggesting that who emerges as the leader of the group matters. Ongoing analyses are also exploring which features of the collaborative discourse are critical for students to take advantage of the affordances of learning by invention.

Keywords: Collaboration, Learning, Problem Solving.

Introduction

It is said that the road to success is paved with failure. It's also said that those who do not learn from their mistakes are destined to repeat them. The provocative implication of these aphorisms is that there may be ways in which failure may be instrumental for successful learning, as long as one is able to take something away from the failure experiences. This is the premise behind learning-by-invention activities. In learning-by-invention, students are asked to attempt to create a mathematical formula to accomplish a goal before an instructional lesson is provided about the canonical approach. The experience of working in a problem space before being told a correct answer may lay the groundwork for future conceptual understanding, and thereby prepare students for future learning. And, these benefits might accrue when solvers become aware of what approaches do not work, or become aware of constraints, obstacles, or desired properties for a solution through previous failures. There is now substantial evidence that having students engage in learning-by-inventions activities in small groups can lead to better understanding of new mathematical and statistical formulas compared to more traditional, direct methods of instruction (e.g., Schwartz & Martin, 2004; Kapur, 2012; Kapur & Bielaczyc, 2011). One main question for our investigations is whether the composition of the small groups in terms of their relative expertise or math ability might affect the likelihood that group members are able to take advantage of learning-by-invention activities. A second main question is whether there are critical features of the group interactions, such as in who emerges as a discourse leader, or what is said during group discussions, that can be shown to facilitate learning.

Although one could expect that groups where all members possess superior math skills would be more successful at any mathematical problem solving activity, another hypothesis is that there may be advantages to being in a group where there are a variety of backgrounds, perspectives or viewpoints that can be contributed. In particular, these investigations are exploring whether diversity in small groups may be one key to unlocking the potential benefits of learning-by-invention activities.

Obviously working in groups with students with more advanced math skills or knowledge may help students with less advanced skills or knowledge by exposing them to advanced math concepts or ideas that they might not consider when working alone. However, it is also possible that collaborating with students with less math knowledge might contribute to more successful problem solving or learning by students with more math knowledge. The work of Webb and others has suggested that more skilled students may benefit from teaching or explaining math concepts to others (e.g. Webb, 1980). In addition, to the extent that students with different mathematical backgrounds might approach problem solving in different ways, then diverse groups may help all members think about a broader range of possible solution approaches which may be particularly important when creative, inventive or innovative thinking is required (Canham, Wiley, & Mayer, 2012; Dunbar, 1995; Wiley & Jensen, 2005; Wiley & Jolly, 2003).

Previous research on learning-by-invention tasks has suggested that groups who generate the widest range of possible solution attempts during the invention phase experience the best learning from the activity (Kapur & Bielaczyc, 2011). Based on this, we predict that the composition of the group in terms of their math expertise should matter, and that there may be special affordances to working in diverse groups. In addition, for groups whose members demonstrate the best understanding of the new principle following learning-by-invention, we explore what features of their discussion may have contributed to their success.

Consistency in Tea

Small groups of undergraduates worked together on an invention activity before receiving a lesson on the standard deviation formula. For the invention activity, students were given data sets representing yearly antioxidant levels for 5-6 years of tea grown by three tea growers. Students were told that a company wished to buy tea from the grower with the most consistent levels of antioxidants from year to year and the company has asked for the students' help. They were prompted to generate as many invented formulas or step-by-step instructions as they could for how they could compute the consistency of antioxidant levels for each tea grower.

Methodology

Two populations of undergraduate Psychology students at the University of Illinois at Chicago participated in this study[1]. Complete data are available for 25 triads who participated as part of a Research Methods course, and 20 triads from an Introductory Psychology course.

Math ability/expertise was estimated based on a median-split derived from historical data from this student population. Students with Math ACT scores of 24 or below were considered to have relatively low skill, and those with scores of 25 or above were considered to have relatively high skill. A score of 25 puts students in the 80th percentile in national norms. Students categorized as having low (M=21.1, SD=2.91) versus high math skill (M=28.5 SD=2.78) differed significantly on the Math ACT, $t(122) = 14.46$, $p < .001$. Of the 45 groups, all students were considered to have low math skill in 11 groups, all students were considered to have high skill in 9 groups, and 25 groups had a mix of high and low skill members. Students were not informed about the nature of their group compositions.

Groups first worked on the invention task for 30 minutes, and discussions were video recorded for the Introductory Psychology groups. For all groups, the worksheets from the invention activity were collected. Following the invention activity, participants individually read through an overview of the standard deviation formula and a worked example of how to compute standard deviation. Following instruction, all students were given a quiz to assess learning outcomes. Two items asked students to apply the formula of standard deviation to a new problem about the weather, and a third item required them to use standard deviation to invent standardized scores in order to compare two students' test performances across different courses. Each item asked students to explain the mathematical reasoning behind their answers. This quiz served as the assessment of learning outcomes from the activity.

Solution Attempts and Quiz Performance

Coding The group worksheets from the invention activity (and video protocols when available) were coded for their inclusion of a variety of different solution approaches to the problem. An initial coding scheme was developed based on categories used by Kapur (2012). It included 5 main categories: 1) computing central tendencies and sums, 2) graphical representations, 3) frequency counts, 4) computing differences between adjacent scores, and 5) computing ranges and deviations from the mean. The final coding scheme with 22 subcategories was established *post hoc* based on an examination of the solutions that were actually obtained so that each distinct solution type had its own subcategory. To score the data, coders assigned each solution attempt to one of the 22 codes. They then determined whether an instance of each subcategory was represented in the written artifacts or not using 0, 1 coding. The total number of different solution approaches was computed by adding the number of subcategories that had at least one instance present in the worksheet or discussion (i.e., the total of the 0, 1 coding across the 22 categories).

In addition, a task analysis of understanding the standard deviation formula was used to identify several critical insights that students might reach during their discussions. The first is that methods such as noticing a high score, graphing histograms or bar graphs, summing scores or computing central tendencies will not help or are not sufficient to quantify consistency. Noticing differences in the range of values across data sets is an important first step

[1] Participants from the Research Methods Sample, who were more advanced in their studies, were found to outperform the Introductory Psychology Sample on the quiz, $F(1,125) = 5.90$, $p < .02$, $\eta^2 = .05$. Importantly, this did not interact with the group composition factor, $F < 1.07$, which meant the two samples could be collapsed in order to increase power, while the sample variable was retained as a covariate in all aggregated analyses reported below (for analyses of samples separately, see Wiedmann, Leach, Rummel, & Wiley, 2012).

toward understanding variance. Two other key insights are that variance is best computed in relation to some reference point (such as a mean), and that somehow variations in positive and negative directions need to be preserved so that they do not cancel out when summed. Based on this analysis, solution attempts that included recognition of range, deviations from the mean, and the need to consider absolute values were all categorized as being of higher quality, and a subtotal of higher quality solution approaches was computed in addition to the overall total number of different solution approaches.

Quiz responses were scored by categorizing each explanation according to the mathematical concept that was referenced. The same basic categories were applied across the 3 problems. Explanations that focused on central tendencies, sums, or maximum scores earned 1 point. Explanations that focused on ranges or differences between scores earned 2 points. If explanations included an incorrect approximation of the SD formula, they received 3 points. If explanations demonstrated a correct use of the SD formula they received 4 points. Combining across the three items, a maximum of 12 points could be reached and the final explanation quality composite score was computed as a proportion of that total. Cronbach's α among the three quiz items was .80. Krippendorff's α indicated good interrater reliability on all coding metrics (> .77).

Quiz Performance An ANCOVA showed a significant effect of group composition on quiz performance, $F(2,123) = 12.41$, $p < .01$, $\eta^2 = .17$. Planned comparisons indicated that students in the all-low math groups had lower scores on the quizzes than students in either the mixed or all-high groups, who did not differ in quiz performance.

A follow-up analysis was performed to see if group heterogeneity affected low-skill and high-skill students differently. As shown in Figure 1, both high- and low-skill members seemed to benefit from participation in mixed groups. A 2x2 ANCOVA (Math Skill x Group Heterogeneity) revealed two significant main effects. As might be expected, high skill students did better than low skill students, $F(1, 122) = 28.44$, $p < .01$, $\eta^2 = .19$. In addition, the main effect for group heterogeneity, $F(1, 122) = 6.29$, $p = .01$, $\eta^2 = .05$, and the lack of a significant interaction, $F < 1$, indicated that both high-skill and low-skill students benefited from working in heterogeneous (mixed) groups.

Solution Attempts Average totals of different solution approaches as a function of group composition are shown in Figure 2. Examples of the inscriptions made on worksheets during the different kinds of solution attempts are shown in Figure 3. An ANCOVA on the total number of different solution approaches showed a significant effect of group composition, $F(2, 41) = 8.55$, $p = .001$, $\eta^2 = .29$. Planned comparisons indicated that the mixed groups considered significantly more different solution approaches than the all-low and all-high groups, who did not differ.

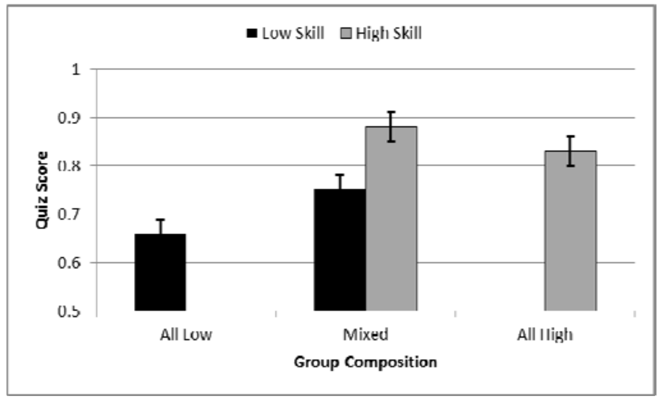

Figure 1: Quiz performance by group composition

When only higher quality solution approaches were considered, a different pattern emerged. An ANCOVA on the number of higher quality representations included in the group worksheets showed a significant effect of group composition, $F(2, 41) = 9.47$, $p < .001$, $\eta^2 = .32$. Planned comparisons indicated that the all-low groups considered fewer different high-quality solution approaches than the all-high and mixed groups, who did not differ. Although the mixed groups also tended to include higher numbers of low-quality solution approaches, this effect did not reach significance, $F(2, 41) = 2.76$, $p < .08$, $\eta^2 = .12$.

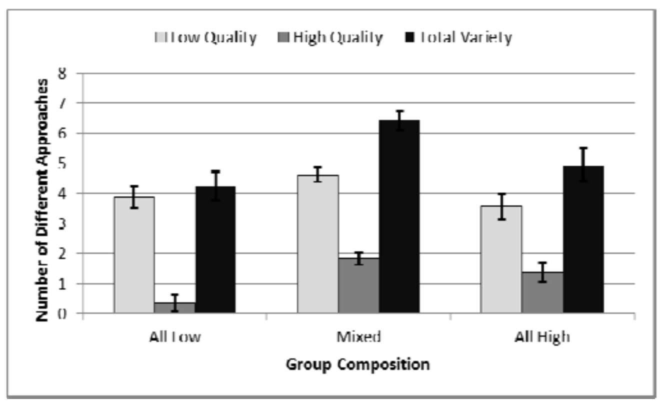

Figure 2: Solution attempts by group composition

Mediational analyses suggested that it was the discussion of a wide range of solution approaches during learning-by-invention activities, including the number of higher quality solution attempts, that mediated the effects of group composition. Heterogeneity predicted the variety of representations, $B = 1.83$ ($SE = .27$), $t(126) = 6.61$, $p < .05$, and variety of representations predicted quiz performance, $B = .02$ ($SE = .01$), $t(126) = 2.47$, $p < .05$. The total effect of heterogeneity on quiz performance was also significant, $B = .09$ ($SE = .03$), $t(126) = 2.84$, $p < .05$. However, this relationship decreased to non-significance when the mediating influence of the variety of representations was included in analysis, $B = .04$ ($SE = .04$), $t(126) = 1.23$, $p = .22$.

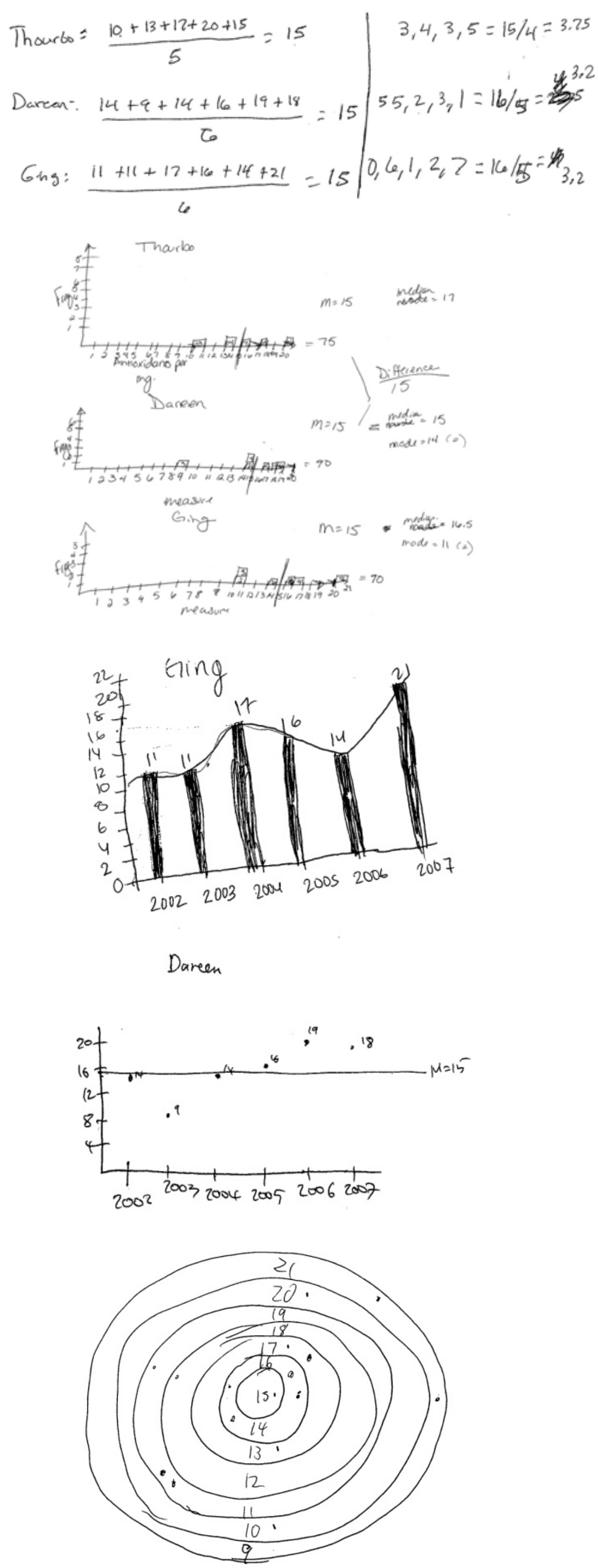

Figure 3: Example inscriptions from solution attempts

In addition, the indirect effect (the mediated effect) of heterogeneity on quiz performance through representation variety was 0.05 (SE = 0.02), and the 95% bias-corrected confidence intervals for the size of the indirect effect did not include zero, (.01, .08) which shows that the indirect effect was significant at a p = .05 level (Preacher & Hayes, 2004, 2008; Shrout & Bolger, 2002). Taken together, these findings provide evidence for full mediation. This analysis suggests that heterogeneity in groups led to better quiz performance because it affected the variety of solutions that were discussed during the learning-by-inventing activity. Additional analyses showed that the benefits of solution diversity during group discussion were demonstrated to contribute to better quiz performance even when the math ability of the students was taken into account.

Analyses of Group Interactions

The second phase of analyseshas been attempting to understand conditions led to the success of the more diverse groups. In this pass through the data, the potential effects of group interactions, such as who emerges as a discussion leader, as well as the quality and content of group discussions, on learning outcomes are examined. Research on decision making groups using hidden profile paradigms has demonstrated the important role of leaders and experts in information sharing. For example, recognition of expertise has been found to be critical for increasing contributions to the group by expert members (Franz & Larson, 2002). In addition, group members are more likely to share valuable information when they are assigned a high status position, such as a group leader (Wittenbaum, Hollingshead, & Botero, 2004). This supports the hypothesis that the expertise of group leaders may be an important predictor of effective information sharing in learning-by-invention tasks that can subsequently influence learning outcomes. Differences in the group discourse and their relation to learning outcomes are also being explored.

Coding Discourse coding was performed on the 14 mixed group discussions for which video recordings were available. Leadership was operationalized by identifying the group member who contributed the largest proportion of utterances to the group discussion. Seven of the groups in the sample had a leader high in math skill, and seven of the groups had a leader low in math skill.

A second goal for the discourse analyses was to explore the content and nature of the discourse acts. Each utterance made by a group member was coded into one of the following categories: (a) solution proposal, (b) clarification request or response, (c) evaluative comment, (d) comment related to group task coordination, (e) calculation, (f) comment on expertise, (g) comment about being stuck or at impasse, or (h) off-task comment.

Leadership When the effects of these two leader types were examined, we found that groups with high math leaders discussed more high quality solution attempts (M = 3.00, SD

= .38) than groups with low math leaders (*M* = 1.57, *SD* = .79), $t(12) = 2.97$, $p < .05$. In terms of learning outcomes, members of groups with high math leaders scored higher on the quiz assessing their understanding of the standard deviation formula (*M* = .82, *SD* = .16) than members of groups with low math leaders (*M* = .68, *SD* = .16), $t(40) = 2.78$, $p < .01$. The results of a 2x2 between-subjects ANOVA (math skill by leader type) with quiz performance as the dependent variable revealed no significant interaction ($F(1, 38) < 1$), suggesting that the expertise of the leader benefited both high and low skill students similarly.

Content of Discussions On average, these groups contributed around 180 utterances during the invention discussion. Only about 10 of these utterances were proposals or amendments to proposals for solution methods. Almost half of the comments were clarifications or requests for clarifications about a proposed solution. About 30 were evaluations of suggested approaches. All other categories represented 10% or less of the utterances.

Results from this discourse analysis suggest that groups who generated a wider range of solutions, proposed more solutions, made more clarifications of proposed solutions, and made fewer comments about being at impasse. A more interesting observation is that they also engaged in more discussion of task coordination. When explored in the context of leader type, groups with a high math leader made more comments related to task coordination, fewer comments about being at impasse, and devoted less discussion to determination of math expertise. Ongoing analyses are more specifically examining who contributes what to the discussion and when. Analyses suggest that low math students are the ones more likely to make comments about expertise, and also, surprisingly, that they are the ones more likely to contribute evaluations of the proposed solutions. This may account for why evaluative comments do not seem to relate to better performance in this sample. As the discourse analysis deepens and matures, this approach is hoped to generate a better understanding of what features of discussion may be critical for learning from invention, so that these features may be used to engineer the design of effective classroom invention activities.

Discussion

The results of this research have shown that groups with members of different backgrounds or expertise may generate a broader range of solution approaches during invention tasks, and that this may benefit understanding of the canonical solution. Group composition in terms of math skill affects when students are able to get the most out of mathematical learning-by-invention activities. Students who worked in mixed groups were better at explaining their understanding of standard deviation on a quiz following the activity than students who worked in more homogeneous groups. Significant effects of group composition were seen in the variety of solution approaches that were considered by groups, particularly higher quality approaches. Interestingly, it was the mixed groups who generated the widest variety of solution attempts, suggesting that they seem to be in a particularly good position to make the most of invention exercises. This is consistent with several other findings suggesting that diversity in expertise among group members contributes to more adaptive, flexible and creative problem solving (Canham, Wiley, & Mayer, 2012; Goldenberg & Wiley, 2011; Wiley & Jensen, 2005; Wiley & Jolly, 2003). Additionally, the consideration of a wider variety of solution approaches during the invention phase, including a larger number of higher quality approaches, predicted the uptake of a later lesson about the standard deviation formula and mediated the effects of group composition and diversity on learning.

To further explore the conditions that might enable effective learning from invention, we found that who emerges as the leader of a diverse group matters. Mixed groups with high math leaders discussed more high quality solution attempts as compared to groups with low math leaders. Interestingly, our discourse coding is also suggesting an important role for defining or coordinating the task among group members (c.f. Moreland *&* Levine, 1992). Groups with high math leaders made more comments in relation to task execution which seemed to relate to their productivity. Yet, it is important to recognize that the leaders self-selected in this study and this can introduce many reasons why these particular groups may have been more or less effective. We are currently conducting a follow-up experiment, again in the context of statistics instruction as part of an undergraduate Psychology course, where we will be assigning high math and low math students to be leaders for the small group activity. Experimental assignment is critical for determining whether and how the expertise of the group leader itself may be important for effective learning from invention activities.

The results thus far suggest that generating a wide variety of approaches to solution may be one important factor determining whether invention discussions prepare students for later learning. Yet, in some cases a richer discussion around fewer alternatives may also lead to successful learning-from-invention, especially if the discussion leads to key insights. Alternatively, we have some evidence that a few of the groups seemed to benefit from the visual affordances of the graphical representations they made. It is possible that some specific kinds of solution attempts may be particularly helpful toward preparation for future learning (i.e., more visual ones or more abstract ones; Ainsworth, 2006, Schwartz, 1995).

The continued analysis of the discussion protocols is intended to serve as source of insight on what particular behaviors one may wish to support while students engage in learning-by-invention tasks. Thus far interactions among group members have not been scripted, roles have not been assigned, and students have not been given any specific direction how to engage in the task together. A next step that others have already begun pursuing (Kapur & Bielaczyc, 2011; Roll, Aleven & Koedinger, 2009) is to provide some support to students in order to maximize the benefits of engaging in invention tasks, but not so much support that the benefits of invention over direct instruction are nullified. Indeed, in most of Webb's previous studies showing benefits of peer collaboration on learning in math, the peer interaction was carefully scaffolded which may have allowed for more stable benefits of mixed ability groups to emerge. One goal for the closer analysis of our discussion protocols is to gain an even better understanding of the conditions that facilitate learning by invention, and how we can capitalize on the intriguing possibility that exploration and failure can sometimes reap benefits toward more sophisticated conceptual understanding in mathematics and statistics.

Acknowledgments

This work has been supported by a Humboldt Research Fellowship to the first author, supporting her research stays in Germany, and a Studienstiftung des Deutschen Volkes Fellowship to Michael Wiedmann, supporting his research stay in the US. A portion of the work reported here served as part of a thesis project submitted by Michael Wiedmann in partial fulfillment of the requirements for the Diploma in Psychology at the University of Freiburg. The authors thank Matthew Cavers, Kelly Currier, Pat Cushen, Thomas Griffin, Robert Hickson, Sabrin Jaber, Allison Jaeger and Ryan Leach for their assistance with data collection, coding, and their diverse contributions to collaborative discussions.

References

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction, 16*, 183-198.

Canham, M., Wiley, J., & Mayer, R. (2012). When diversity in training improves dyadic problem solving. *Applied Cognitive Psychology, 26*, 421-430.

Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 365-395). Cambridge, MA: MIT Press.

Franz, T. M., & Larson, J. R., J. (2002). The impact of experts on information sharing during group discussion. *Small Group Research, 33*, 383-411.

Goldenberg, O., & Wiley, J. (2011). Quality, conformity, and conflict: Questioning the assumptions of Osborn's Brainstorming technique. *Journal of Problem Solving, 3*(2).

Kapur, M. (2012). Productive failure in learning the concept of variance. *Instructional Science, 40*, 651-672.

Kapur, M., & Bielaczyc, K. (2011). Designing for productive failure. *The Journal of the Learning Sciences, 41*, 45-83.

Moreland, R. L., & Levine, J. M. (1992). Problem identification by groups. In S. Worchel, W. Wood, & J. A. Simpson, (Eds.), *Group processes and productivity.* Newbury Park, CA: Sage.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, and Computers, 36*, 717-731.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*, 879-891.

Roll, I., Aleven, V., & Koedinger, K. R. (2009). Helping students know 'further' - increasing the flexibility of students' knowledge using symbolic invention tasks. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, 1169-1174.

Schwartz, D. L. (1995). The emergence of abstract representations in dyad problem solving. *Journal of the Learning Sciences, 1*, 321-354.

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*, 129–184.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7,* 422-445.

Webb, N. M. (1980). A process-outcome analysis of learning in group and individual settings. *Educational Psychologist, 15*, 69-83.

Wiedmann, M., Leach, R., Rummel, N. & Wiley, J. (2012). Does group composition affect learning by invention? *Instructional Science, 40*, 711-740.

Wiley, J. (1998) Expertise as mental set: The effects of domain knowledge in creative problem solving. *Memory & Cognition, 26*, 716-730.

Wiley, J., & Jensen, M. (2006). When three heads are better than two. *Proceedings of the 28th Annual Conference of the Cognitive Science Society,* 2375-2380.

Wiley, J. & Jolly, C. (2003). When two heads are better than one expert. *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society,* 2375-2380.

Wittenbaum, G. M., Hollingshead, A. B., & Botero, I. C. (2004). From cooperative to motivated information sharing in groups: Moving beyond the hidden profile paradigm. *Communication Monographs, 71*, 886-310.