

Chapter 14

Mathematical Skills and Learning by Invention in Small Groups

Michael Wiedmann, Ryan C. Leach, Nikol Rummel, and Jennifer Wiley

Abstract The purpose of the present research was to investigate how the effectiveness of learning-by-invention activities may be influenced by the composition of the small groups that engage in them in terms of the mathematical skills of their members. Undergraduates engaged in an “inventing standard deviation” activity. Groups that included both high- and low-skill members generated a broader range of solution attempts and more high-quality solution attempts during the activity. Both the range and quality of solution attempts that were generated related to better uptake of the standard deviation formula from a later lesson. These results suggest that the composition of the small groups that work together may have an impact on the effectiveness of learning-by-invention activities.

Keywords Collaborative learning • Collaborative problem solving • Learning by invention • Group composition • Mathematical skill

Mathematical Skills and Learning by Invention in Small Groups

One approach for teaching new mathematical procedures is to provide direct instruction with a lesson that introduces the new problem-solving method (c.f. Anderson et al. 1995; Rosenshine and Stevens 1986). After the lesson, students are encouraged to practice using the new formula. This approach makes sure that students have the prior knowledge necessary to solve problems with the new formula. But does presenting the lesson first lead to the best understanding of the formula? Or are students

M. Wiedmann • N. Rummel
Institute of Educational Research, Ruhr-Universität Bochum, Bochum, Germany

R.C. Leach • J. Wiley (✉)
Department of Psychology, University of Illinois, Chicago, USA
e-mail: jwiley@uic.edu

able to come up with useful attempts for solving new problems on their own? In an alternative approach to mathematics instruction, *learning by invention*,¹ students (generally working in small groups) attempt to invent their own solution methods before being taught the canonical formula. Learning by invention has now repeatedly been shown to be just as effective as instruction where the canonical solution is taught first (e.g., Belenky and Nokes-Malach 2012; DeCaro and Rittle-Johnson 2012; Kapur 2009, 2012; Kapur and Bielaczyc 2011; Loibl and Rummel 2013; Roll et al. 2009; Schwartz and Martin 2004; Westermann and Rummel 2012). This chapter takes a closer look at the question of whether diversity in math skills among the members of the small groups might play a role in learning by invention.

As an example of the efficacy of learning by invention for teaching statistics, Schwartz and Martin (2004) compared two instructional conditions. In the experimental condition, students engaged in learning by invention. Their invention task was to compare data across different distributions in order to develop standardized scores. Following the invention activity, students then received a worked example teaching them about this concept. Thus, the learning-by-invention condition involved both an invention phase and an instruction phase. In a comparison condition, students were first taught how to standardize scores before practicing the procedure. Students in the learning-by-invention condition outperformed the control condition on a transfer test that required applying standardized scores in a new context. Schwartz and Martin argued that the invention phase served as *preparation for future learning* from the worked example. The invention process was suggested to activate prior knowledge that facilitated learning from the following direct instruction. The creation and careful consideration of solution attempts during the invention phase may be a mediator of this effect.

Kapur (2009, 2012) has also shown benefits from learning by invention on conceptual and procedural learning. In his work, he proposes that invention may lead to productive failure; that is, students may fail at generating a formula, but, similarly to Schwartz and Martin (2004), that this will be productive for future learning. Kapur's instructional approach combines the invention phase with a class lecture and discussion in which students' solution attempts are compared and contrasted with each other and with the canonical solution. This may help students recognize the critical constraints and affordances of these solutions. Kapur (2012) showed that, non-surprisingly, students in an invention condition generated a more diverse set of solutions than the control condition that was taught canonical solutions. However, students in the invention condition also outperformed the control condition on conceptual understanding items and performed just as well as the control

¹In the literature, there are several similar approaches that explore the learning opportunities that can result from having students engage in problem solving before having received instruction. VanLehn referred to this as *impasse-driven learning* (1988). Schwartz and Martin (2004) have argued that an initial invention phase can provide "preparation for future learning." Kapur frames generation and exploration (Kapur 2012), or elicitation phases (Kapur and Bielaczyc 2011), as ways to encourage *productive failure*. In the present paper, we adopt the terminology and instructional sequence of Schwartz and Martin (2004) where an invention phase is followed by instructional support in the form of a worked example.

condition on procedural fluency items. In addition, Kapur and Bielaczyc (2011) also found that the diversity of solutions generated during invention predicted posttest performance. This empirical link between the invention process and learning suggests that the consideration of multiple solution approaches during invention activities may be one key to students' preparation for future learning.

The Present Study

The previously reported results suggest that learning by invention may be an effective instructional approach for promoting conceptual understanding of formulas as well as procedural knowledge of how to use them (following Mayer and Greeno 1972). The main question of the present study is *under which conditions* this approach may be most effective. In particular, we are investigating how the small groups engaging in invention activities may be best composed in order to optimally support individual student learning. Since the problem-solving task is mathematical, it seems likely that the mathematical skill of individual group members may have an effect on group interaction. Roll (2009) was only able to show benefits from invention activities in high school students that took college-level (Advanced Placement) courses, but not for more typical students. Kapur and Bielaczyc (2011) investigated benefits of productive failure in three schools of varying student profiles in mathematical skill. The effect of productive failure activities was stronger in schools of higher mathematical skill profiles. Therefore, one prediction might be that learning by invention is only effective for students with higher mathematical skill.

However, it may also be sufficient that *each group* has at least one member with higher mathematical skill (Wiley et al. 2009). Many researchers (Paulus 2000; Strobe and Diehl 1994; Wiley and Jensen 2006; Wiley and Jolly 2003) have suggested that diversity in the background of group members may be beneficial for problem solving. Dunbar (1995) showed that in laboratories where scientists came from different disciplines, unexpected findings led to many more alternate hypotheses and analogies, which in turn led to more scientific breakthroughs. Gijlers and de Jong (2005) found that dyads engaging in discovery learning generated more hypotheses when they were heterogeneous in prior knowledge than when they were homogeneous. And Canham et al. (2012) found that dyads were better at solving transfer items when their members were trained in different ways of solving probability problems than when both members had received the same training.

Heterogeneous group composition in terms of the math skills of the members may also influence the interaction of the group. Webb (1980), for instance, found that when high- and low-skill students work together, they often form teacher-student relationships. This peer tutoring can not only be beneficial for the tutee but also for the high-skill tutor. Webb also found that working in mixed groups seemed to promote the most explanation-giving during group discussion. Given these advantages of heterogeneous group composition, it may also be that in invention activities, mixed groups will have the most productive discussions. However, it is

also possible that the high-skill members will show poorer learning outcomes when having to work with low-skill students than when working in homogeneous, high-skill groups (Fuchs et al. 1998). It is therefore an interesting question whether mathematical skill of each group member, and group composition in terms of mathematical skill of the members, may have an effect on learning by invention.

To test whether the composition of groups in terms of their math skills might matter, the present study explored differences in the effects of learning by invention in performance among three group types: all-low-skill groups, all-high-skill groups, and mixed groups. The target content was the standard deviation formula, and mathematical skill was measured using scores on a standardized college admission test (the Math ACT). Data was collected in two contexts. Some groups participated as part of an undergraduate course in Research Methods in Psychology. For these students, dependent measures included written artifacts of the invention process and an online quiz to assess learning. A second sample was collected from a subject pool of undergraduates enrolled in Introduction to Psychology. These students participated in a laboratory study using parallel procedures, but additionally recordings were collected that allowed for a more complete accounting of the group discussion.

The main hypotheses to be tested were (1) whether groups needed at least one high math member to take advantage of learning by invention and (2) whether heterogeneous group composition (i.e., participating in mixed groups) would positively affect the variety and quality of solution approaches generated during the invention activity, which would in turn affect learning. Thus, the main analyses of interest were ANOVAs testing for the main effect of group composition on both solution variety and quiz performance, with planned comparisons among the three different group types. Subsequent analyses tested whether solution variety and quality would predict quiz scores, acting to mediate the effect of group composition on performance.

Method

Participants

Research Methods Sample

Students who enrolled in an undergraduate Research Methods course in Psychology at the University of Illinois at Chicago participated in the experiment as a class activity. This course is usually taken in the second year of university. Students who take this course generally intend to declare psychology as their major.

The original sample consisted of 149 students, taught in six sections and assigned to groups of three based on their Math ACT scores so that there would be groups in each category of group type. Students were unaware that ACT scores were used to

assign them to groups. Assigning students to groups also prevented established groups from working together, to make this study more similar to the randomly assigned groups obtained in the subject pool sample. Students had to be excluded for several reasons: Because Math ACT scores were not available for all students, 66 students from groups where some members' Math ACT scores were unknown were excluded from both group-level and individual-level data analyses. Another 15 students did not complete the final quiz. Those students, but not the other members of their groups, were excluded from learning outcome analyses resulting in a final sample size of 68 individuals for individual-level analyses. There was data from members of 25 groups available for group-level analyses.

Participants received credit for participating in the activity and completing the homework assignment, as they did for all recitation and homework activities in their class. They were unaware that the quiz would not count toward their grade. The homework assignment, which included the quiz, was announced after the invention activity.

Introduction to Psychology Sample

Sixty undergraduate students from the Introduction to Psychology course at the University of Illinois at Chicago were recruited to participate in the experiment as part of a subject pool. Introduction to Psychology is typically taken during the first or second semester of university. Groups were comprised of students who signed up individually for the same time slot. Skill profiles of the groups were ascertained after the data was collected. Groups of friends who signed up together were excluded from further analysis. There were 59 students with complete data that could be included in the individual analyses, and data from members of 20 groups were available for group-level analyses.

Math Skill Level

For both samples, math skill level was 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 lower skill, and those with scores of 25 or above were considered to have higher skill. A score of 25 puts students in the 80th percentile in national norms. Of the 127 individuals available for individual analyses, 64 were classified as low math skill and 60 as high math skill. Students categorized as having high versus low math skill 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.

Materials

Invention Activity

The invention activity used in this study is included in Appendix A of Wiedmann et al. (2012). This activity was based on prior invention activities developed by Kapur (2012) and Schwartz and Martin (2004) in which students are tasked with comparing three data sets. In this study, the invention activity used a cover story about the amount of antioxidants found in tea coming from 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 are asked to propose a formula for calculating the consistency of antioxidant levels for each tea grower.

Quiz

The quiz contained three items: two in which the formula for standard deviation needed to be applied to a new problem about the weather and one in which students needed to invent standardized scores in order to compare two students’ test performances across different courses. Students were asked to explain the mathematical reasoning behind their answers. This quiz served as the assessment of learning outcomes for the activity and is based on items used in Kapur (2012).

Procedure

Research Methods Sample

The study took place as part of a course in Research Methods, during the weekly recitation section meeting. At the start of the meeting, the teaching assistant gave a short (10 min) introduction that began with an example research question and two data sets. For each data set, the teaching assistant demonstrated how to draw a histogram and defined and calculated the mean and median. While the means were the same in both data sets, the medians were not. To help the student notice the variance among scores, students were then asked to describe the other big difference they could see between the two data sets.

Students then worked in groups for 30 min with the goal of inventing a formula to describe “consistency” in three data sets.

They were given a group worksheet with three data sets. The worksheet asked them to generate as many invented formulas as they could to describe consistency in the three data sets, and provided additional space for their solution attempts. The group worksheets were collected at the end of the discussion.

After class, students completed an online homework assignment through the university's e-learning (Blackboard) system. As usual, they completed the homework individually at a time of their choosing before the next class meeting. This assignment included a short lesson about the standard deviation formula and asked students to compute standard deviations from a worked example before the quiz (following Schwartz and Martin 2004).

Introduction to Psychology Sample

The procedure was largely the same, except that the small groups were run one group at a time in a laboratory room and recorded. The introduction given by the experimenter was similar except median was not mentioned. Because students sometimes are overwhelmed with the demand to create a formula (Roll et al. 2009), in this sample, it was clarified that instead of a formula, they could also write step-by-step instructions for how they would compute consistency.

The remainder of the procedure was similar. After working on the invention activity together for 30 min, the group members were separated to work individually for the remainder of the study. Each student was given the overview of the standard deviation formula and worked example to read before taking the quiz.

Coding Schemes

Coding of Solution Attempts

The group worksheets from the invention activity were coded for both variety and quality of solutions. A coding scheme was established post hoc based on the range of solutions that were actually obtained such that each distinct solution type had its own subcategory. A list of the 22 final codes appears in Appendix B of Wiedmann et al. (2012). Coders assigned each solution attempt to one of the 22 subcategories. The total number of different solution approaches was computed for each group by adding the number of subcategories that had at least one instance present in the group worksheet (i.e., the total of the 0, 1 codings across the 22 codes).

To code for differences in quality of solution attempts, a task analysis of understanding the standard deviation formula identified several critical insights that students might reach during their discussions. The first insight is that methods such as making histograms or bar graphs, noticing an individual high or low score, or summing or averaging scores will not help to quantify consistency. Alternatively, noticing differences in the range of values across data sets is an important first step toward understanding variance. A second key insight is that somehow variations in positive and negative directions need to be handled in some way so that they do not cancel each other out. A third key insight is that variance needs to be computed in relation to some reference point (such as the mean). 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 variety of solution approaches.

Coding for the Research Methods sample relied on the worksheets. Coding for the Introduction to Psychology sample was also based on ideas mentioned in discussion when transcripts of the discussions were available. Two individuals coded all groups for the presence or absence of solution attempts in each subcategory (Krippendorff $\alpha = .81$). Differences were resolved by a third rater.

Coding for Quiz Responses

Each of the three problems was scored using the same basic concepts and point values, giving the student the point value assigned to the most advanced concept that was referenced in each explanation:

Central tendency, sum, or maximum score (1 point)

Examples: The average of February is higher than January, so they should go with January. Alicia was only 1 point away from a perfect score. Alicia had a higher score.

Ranges and deviations: differences between scores, subtracting smallest from largest score, differences from the mean (2 points)

Examples: The difference from the temperature for February by month is 2, 2, 1, 3, 4 and that is very consistent. January has a lower range. Chemistry has more of a spread. Alicia is further from the mean.

Vague or incorrect formula or reasoning about *SD* (3 points)

Examples: A higher deviation means the classes were harder, making Alicia more deserving.

Correct use of *SD* (4 points)

Examples: January has a lower standard deviation. Kelvin should receive the award because his score has a greater number of standard deviations above the average.

Two individuals scored all posttest items. A maximum score of 12 points was possible across the 3 items. 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 three items (item 1 = .84, item 2 = .81, item 3 = .77).

Results

Learning Outcomes

Before proceeding to test the main questions, we explored the independency of the individual learning data since it was obtained in a group setting. Kenny et al. (1998) suggest the calculation of intra-class correlations to test for consequential nonindependence. Because the intra-class correlation for group members' quiz scores was

not significant in the Research Methods sample, $ICC = .08$, $p = .55$, $CI = 95\%$, and the Introduction to Psychology sample, $ICC = .12$, $p = .36$, $CI = 95\%$, it was appropriate to analyze learning outcomes on an individual level.

In a next step, differences between the two samples were explored. Participants from the Research Methods sample, who were more advanced in their studies, were found to outperform the Introduction to 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 more complete analyses of this data, including descriptive statistics and analyses for the separate samples, see Wiedmann et al. 2012).

The top panel of Fig. 14.1 presents average quiz performance as a function of group composition (entered as a nominal variable) and math skill. An ANCOVA with sample entered as a covariate showed a significant effect of group composition

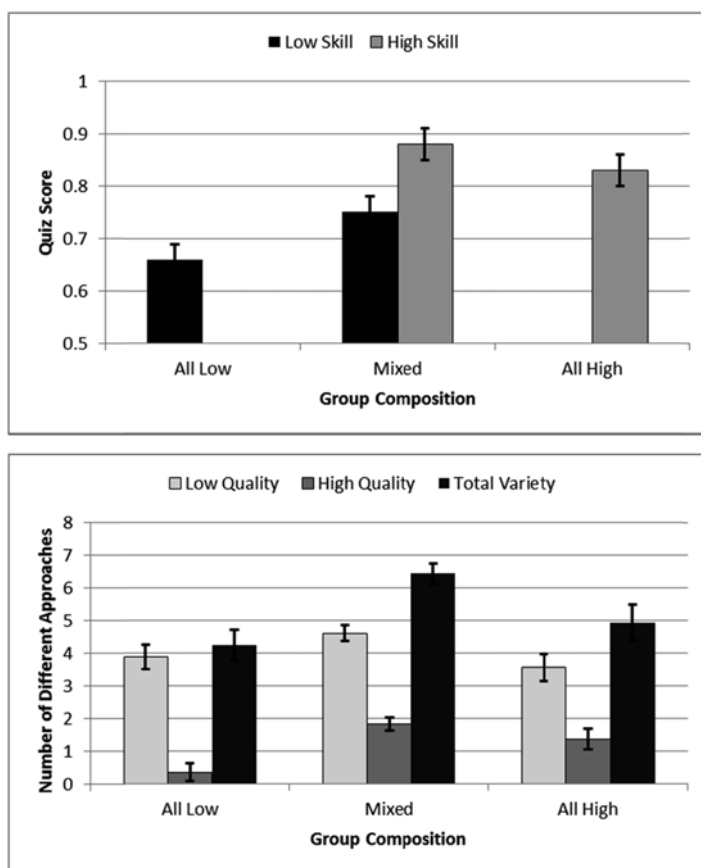


Fig. 14.1 Adjusted means for quiz scores (*top*) and solution variety (*bottom*) by group composition

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 the top panel of Fig. 14.1, both high- and low-skill members seemed to benefit from participation in mixed groups. A 2×2 ANCOVA (math skill \times group heterogeneity) with sample entered as a covariate 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.

Variety of Solution Approaches

Average totals of different solution approaches as a function of group composition are shown in the bottom panel of Fig. 14.1. An ANCOVA on the total number of different solution approaches with sample entered as a covariate showed a significant effect of group composition, $F(2, 41)=8.55, p = .001, \eta^2 = .29$. Planned comparisons indicated that the mixed groups documented significantly more different solution approaches than the all-low-skill, $p < .001$, and all-high-skill groups, $p = .02$, who did not differ, $p = .33$.

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 documented fewer different high-quality solution approaches than the all-high, $p = .02$, and mixed groups, $p < .001$, who did not differ, $p = .23$. Although the mixed groups also tended to include higher numbers of low-quality solution approaches, this effect did not reach significance, $F(2, 41)=2.76, p < .08, \eta^2 = .12$.

Relation of Solution Variety to Learning Outcomes

The partial correlations among the total number of different solution approaches, high-quality solution approaches, low-quality approaches, and students' quiz scores (controlling for sample) are presented in Table 14.1.

Two final analyses were then performed to test whether the discussion of a broad variety of representations was responsible for the better performance that was observed as a function of group heterogeneity. To investigate this mediational hypothesis, the test of indirect effect procedure and corresponding macro (Preacher

Table 14.1 Correlations between number of solutions and quiz performance

	Low quality	High quality	Total variety
Quiz score	.15	.34**	.31**
Low quality		.14	.81**
High quality			.70**

Note: $N = 127$, $df = 124$, ** $p < .01$

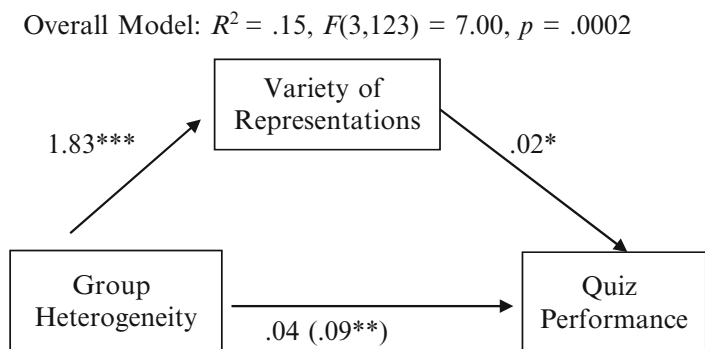


Fig. 14.2 Mediation model. Note: the value in parentheses indicates the total effect before accounting for mediation. * $p < .05$, ** $p < .01$, *** $p < .001$

and Hayes 2008) was employed using 5,000 resamples. For these analyses, bootstrapping tests are generally preferred to the more traditional Sobel test because they do not assume a normal distribution of the product terms which are usually normally distributed only in large samples (Preacher and Hayes 2004, 2008; Shrout and Bolger 2002). Mixed groups were coded as “1” for heterogeneity, and the remaining groups were coded as “0” for this analysis. Results indicated that heterogeneity predicted the variety of representations, $B = 1.83$ ($SE = .27$), $t(126) = 6.61$, $p < .05$, and that 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 the analysis, $B = .04$ ($SE = .04$), $t(126) = 1.23$, $p = .22$ (see Fig. 14.2).

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 and Hayes 2004, 2008; Shrout and 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.

Of course, another critical aspect of group composition as defined in this study was that it was based in math skill (all low, mixed, and all high). Clearly math skill can have a direct effect on learning about math for any individual, so it is interesting to ask if the discussion of a variety of representations during the learning-by-inventing activity might have a significant effect on performance unique from the effect of group composition on performance through math skill.

To address this concern, we performed a second regression analysis using both representation variety and math ACT scores as mediators. This analysis showed that group composition significantly predicted the variety of representations produced by the group, $B = .56$ ($SE = .23$), $t(123) = 2.45$, $p < .05$, and math ACT scores, being the basis upon which group composition was defined, were significantly related to composition, $B = 3.72$ ($SE = .54$), $t(123) = 6.89$, $p < .05$. Both representation variety, $B = .02$ ($SE = .01$), $t(123) = 2.72$, $p < .05$, and math ACT scores, $B = .02$ ($SE = .00$), $t(123) = 4.76$, $p < .05$, significantly predicted quiz performance. The total effect of group composition on quiz performance was also significant, $B = .10$ ($SE = .02$), $t(123) = 4.33$, $p < .05$, but was reduced to non-significance when including the mediating influences of representation variety and math skill, $B = .03$ ($SE = .02$), $t(123) = 1.04$, $p = .30$. The indirect effect through variety of representations was 0.01 ($SE = 0.01$), and importantly, the 95% bias-corrected confidence intervals for the size of the indirect effect did not include zero (.003, .03). Taken together, these results indicate full mediation by variety of representations even when the effects of math skill are included in the analysis.

When these same two mediational analyses were performed using the number of high-quality solutions instead of total variety measures, identical patterns of results were found. The discussion of more high-quality solution approaches also mediated the group homogeneity and composition effects and contributed to performance independently of math ACT scores.

Taken together, these mediational analyses suggest that it is the discussion of a wide range of solution approaches during learning-by-invention activities (including a number of higher-quality solution attempts) that mediates the effects of group composition. More diverse groups documented a broader variety of solution approaches, and when more solution approaches were documented, that improved performance on later quizzes. Further, the benefits of solution diversity during group discussion were demonstrated to contribute to a better quiz performance even when the math skill of the students was taken into account.

Discussion

The results of this study suggest that group composition in terms of math skill affects whether students are able to benefit from 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 homogeneous groups. Significant effects of group composition were

seen in both variety and quality of solution 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 result converges with several other findings in suggesting that diversity in expertise among group members can contribute to more adaptive, flexible, and creative problem solving (Canham et al. 2012; Gijlers and De Jong 2005; Goldenberg and Wiley 2011). Additionally, the consideration of a wider variety of solution approaches during the invention phase, including a 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.

These results show a significant benefit of working in mixed groups for learning-by-invention activities. Yet, more research is needed to fully understand the affordances of this instructional context. It is possible that even more robust effects could be found with a longer invention activity, a conjecture that could be explored in future research. The invention activity used here was of a fairly short duration, and a number of the groups seemed to be approaching some critical insights when time ran out (Wiedmann et al. 2012). In previous studies, students generally engaged in their invention discussions for more than one class period (Kapur 2012; Schwartz and Martin 2004).

Another limitation of the present study was the lack of a pretest-posttest design to demonstrate that better quiz scores reflected improved learning from the activity. Also, because the present studies did not include a direct instruction comparison condition, these results cannot speak to whether low-skill students may benefit more from learning by invention in mixed groups than they would have from direct instruction.

One recommendation for future studies would be to consider using an instruction that does not prompt for a formula at all. In a number of groups, arbitrary formulas were contributed during the discussion. These formulas were not attempts to quantify a particular solution approach that was being discussed qualitatively. Instead, students just brought up simple formulas that students knew like $\text{distance} = \text{rate} \times \text{time}$. We suspect this problematic behavior may have been a consequence of giving the instruction “to create a formula” in these studies. It may be better to instruct students to give step-by-step descriptions of how to compute consistency (Roll et al. 2009) or to prompt students to generate a method (Schwartz and Martin 2004). For the Introduction to Psychology sample, we included requests for both formulas and step-by-step descriptions as part of our task instruction; however, many students still seemed to focus on the formula goal.

Because the benefits of learning by invention over direct instruction may be less robust for low-skill students (Kapur and Bielaczyc 2011; Kroesbergen et al. 2004; Roll 2009), all of the above points represent important issues for future research. Further, while these results represent some of the first demonstrations of learning by invention for low-skill students, an important observation is that previous attempts have used much younger samples. We suspect all college students will have the capacity to engage in the demands of this learning-by-inventing task, even if the low-skill students are less proficient at math tests. Given this, it is possible that the present

findings will not generalize to younger samples where the demands of a learning-by-invention activity may present too much of a challenge for low-skill learners. It is an interesting question for future research whether the benefits of working in mixed groups can be seen in younger samples, which would be consistent with other work (i.e., Webb 1980) showing learning benefits when students with different ability levels work together.

Another important direction for this line of research is the further exploration of what is happening during these collaborative discussions that is critical for effective learning from invention. The analyses so far have shown that a broader variety of representations are discussed and a larger number of higher-quality solution attempts are considered, but how are these brought into the conversation? The really interesting questions of how the interactive discourse and dynamics of mixed groups may facilitate learning from invention have yet to be answered.

We have only just begun the task of analyzing the discussion protocols of groups, starting with the three most successful mixed groups of the Introduction to Psychology sample (Wiedmann et al. 2012). Some initial impressions suggest that there are multiple ways in which groups can engage in invention activities. In our preliminary analysis (reported in Wiedmann et al. 2012), we found that the first group discussed fewer solution approaches than the other two groups, but they seemed to engage in discussion on a more conceptual level. They also engaged in more evaluation of the proposals and in more reflection on their progress. On the other hand, the two other groups generated more solution approaches, but this activity seemed to be accompanied by less discussion. A very preliminary speculation could be that generating a wide variety of approaches to the problem may be one important factor. In addition, a richer discussion around fewer alternatives can also lead to successful learning-by-invention activities as seen in the first group, especially if the discussion leads to key insights. Alternatively, two of the three groups seemed to benefit from the visual affordances of line graphs. 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). Although no universal pattern could be identified for the most successful groups, future analyses exploring the interaction patterns among the least successful groups could reveal more consistency in the behaviors that lead to ineffective collaboration. Other questions for future analyses include: What role do behaviors such as question-asking, responsiveness, evaluating proposals, connecting across representations, and generating or hearing explanations play in group success? How are high-quality approaches being discussed or discovered? What contributions do the high-skill versus the low-skill members make to the discussions? Who is acting as the group leader and how do they lead the group? Other preliminary analyses of the discussion suggest that being in a group with a high-skill leader is critical (Wiley et al. 2013).

Although we have motivated our study by focusing on the contribution of mathematical knowledge by high-skill members, there are other mechanisms by which they may have influenced the groups. For example, invention may be a novel type of exercise for many students. High-skill students may be more familiar with

these tasks, or they may be more willing to engage in novel tasks, or they may possess a greater sense of self-efficacy in math which enables them to have a more positive approach to these tasks. Alternatively, the high-skill students may possess superior metacognitive abilities, and with those they may help the groups to monitor and reflect on their progress or regulate their learning and studying activities. Either of these alternative explanations suggests that high-skill members may not be necessarily contributing specific knowledge to the mixed groups, but may be helping the groups via other attributes that are generally correlated with expertise in a domain. A complete analysis of the discussion protocols from the Introduction to Psychology sample is currently underway which will help to address these questions.

This analysis of the discussion protocols will also be a great source of insight on what particular behaviors one may wish to support while students engage in learning-by-invention tasks. In the present study, we did not script the interactions among group members, did not assign roles, and did not give students any specific direction on how to engage in the task together. Others have already begun to test (Kapur and Bielaczyc 2011; Roll et al. 2012, 2009; Westermann and Rummel 2012) if students can be supported in order to maximize the benefits of engaging in invention tasks, without nullifying the benefits of invention over direct instruction. Indeed, peer interaction was carefully scaffolded in most of Webb's previous studies, which may have allowed for more stable benefits of mixed groups to emerge. Our goal for the closer analysis of our discussion protocols is to help to determine whether these candidate behaviors seem to facilitate learning by invention or if there are other characteristics of successful interactions that emerge. The present study has demonstrated that students may benefit most from learning-from-invention activities when working in mixed groups. Future research needs to further explore why and how these benefits are afforded and, importantly, whether providing supports for these affordances can ensure benefits in all groups.

Acknowledgments This research and collaboration was supported by a Studienstiftung des Deutschen Volkes [German National Academic Foundation] Fellowship to Michael Wiedmann and by a Humboldt Research Fellowship to Jennifer Wiley. The data reported here were collected 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 Kelly Currier, Pat Cushen, Olga Goldenberg, Thomas Griffin, Robert Hickson, Allison Jaeger, and Andy Jarosz for their assistance with coding, data collection, and discussions on this project.

References

- Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction, 16*, 183–198. doi:10.1016/j.learninstruc.2006.03.001.
- Anderson, J. R., Corbett, A. T., Koedinger, K., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences, 4*, 167–207. doi:10.1207/s15327809jls0402_2.
- Belenky, D. M., & Nokes-Malach, T. J. (2012). Motivation and transfer: The role of mastery-approach goals in preparation for future learning. *Journal of the Learning Sciences, 21*, 399–432. doi:10.1080/10508406.2011.651232.

- Canham, M., Wiley, J., & Mayer, R. (2012). When diversity in training improves dyadic problem solving. *Applied Cognitive Psychology*, *26*, 421–430. doi:10.1002/acp.1844.
- DeCaro, M. S., & Rittle-Johnson, B. (2012). Exploring mathematics problems prepares children to learn from instruction. *Journal of Experimental Child Psychology*, *113*, 552–568. doi:10.1016/j.jecp.2012.06.009.
- 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.
- Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Karns, K. (1998). High-achieving students' interactions and performance on complex mathematical tasks as a function of homogeneous and heterogeneous pairings. *American Educational Research Journal*, *35*, 227–267. doi:10.3102/00028312035002227.
- Gijlers, H., & De Jong, T. (2005). The relation between prior knowledge and students' collaborative discovery learning processes. *Journal of Research in Science Teaching*, *42*, 264–282. doi:10.1002/tea.20056.
- Goldenberg, O., & Wiley, J. (2011). Quality, conformity, and conflict: Questioning the assumptions of Osborn's Brainstorming technique. *Journal of Problem Solving*, *3*. Retrieved from <http://docs.lib.purdue.edu/jps/vol3/iss2/5>
- Kapur, M. (2009). Productive failure in mathematical problem solving. *Instructional Science*, *38*, 523–550. doi:10.1007/s11251-009-9093-x.
- Kapur, M. (2012). Productive failure in learning the concept of variance. *Instructional Science*, *40*, 651–672. doi:10.1007/s11251-012-9209-6.
- Kapur, M., & Bielaczyc, K. (2011). Designing for productive failure. *The Journal of the Learning Sciences*, *41*, 45–83. doi:10.1080/10508406.2011.591717.
- Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 1, pp. 233–265). New York: Oxford Press.
- Kroesbergen, E. H., Van Luit, J. E. H., & Maas, C. J. (2004). Effectiveness of explicit and constructivist mathematics instruction for low-achieving students in the Netherlands. *The Elementary School Journal*, *104*, 233–251. doi:10.1086/499751.
- Loibl, K., & Rummel, N. (2013). The impact of guidance during problem-solving prior to instruction on students' inventions and learning outcomes. *Instructional Science*, *42*, 305–326. doi:10.1007/s11251-013-9282-5.
- Mayer, R. E., & Greeno, J. G. (1972). Structural differences between learning outcomes produced by different instructional methods. *Journal of Educational Psychology*, *63*, 165–173. doi:10.1037/h0032654.
- Paulus, P. (2000). Groups, teams and creativity: The creative potential of idea generating groups. *International Journal of Applied Psychology*, *49*, 237–262. doi:10.1111/1464-0597.00013.
- 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. doi:10.3758/BF03206553.
- 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. doi:10.3758/BRM.40.3.879.
- Roll, I. (2009). *Structured invention activities to prepare students for future learning: Means, mechanisms, and cognitive processes*. Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh.
- Roll, I., Aleven, V., & Koedinger, K. R. (2009). Helping students know 'further' – Increasing the flexibility of students' knowledge using symbolic invention tasks. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st annual conference of the cognitive science society* (pp. 1169–1174). Austin: Cognitive Science Society.
- Roll, I., Holmes, N. G., Day, J., & Bonn, D. (2012). Evaluating metacognitive scaffolding in guided invention activities. *Instructional Science*, *40*, 691–710. doi:10.1007/s11251-012-9208-7.

- Rosenshine, B., & Stevens, R. (1986). Teaching functions. In M. C. Wittrock (Ed.), *Handbook of research on teaching*. New York: Macmillan.
- Schwartz, D. L. (1995). The emergence of abstract representations in dyad problem solving. *Journal of the Learning Sciences, 1*, 321–354. doi:10.1207/s15327809jls0403_3.
- 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. doi:10.1207/s1532690xci2202_1.
- Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7*, 422–445. doi:10.1037/1082-989X.7.4.422.
- Stroebe, W., & Diehl, M. (1994). Why groups are less effective than their members: On productivity losses in idea-generating groups. *European Review of Social Psychology, 5*, 271–303. doi:10.1080/14792779543000084.
- VanLehn, K. (1988). Toward a theory of impasse-driven learning. In H. Mandl & A. Lesgold (Eds.), *Learning issues for intelligent tutoring systems* (pp. 19–41). New York: Springer.
- Webb, N. M. (1980). A process-outcome analysis of learning in group and individual settings. *Educational Psychologist, 15*, 69–83. doi:10.1080/00461528009529217.
- Westermann, K., & Rummel, N. (2012). Delaying instruction: Evidence from a study in a university relearning setting. *Instructional Science, 40*, 673–689. doi:10.1007/s11251-012-9207-8.
- Wiedmann, M., Leach, R., Rummel, N., & Wiley, J. (2012). Does group composition affect learning by invention? *Instructional Science, 40*, 711–740. doi:10.1007/s11251-012-9204-y.
- Wiley, J., & Jensen, M. (2006). When three heads are better than two. *Proceedings of the twenty-eighth annual conference of the cognitive science society*, Mahwah: Lawrence Erlbaum.
- 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*, Mahwah: Lawrence Erlbaum.
- Wiley, J., Goldenberg, O., Jarosz, A. F., Wiedmann, M., & Rummel, N. (2013). Diversity, collaboration, and learning by invention. *Proceedings of the 35th annual meeting of the cognitive science society*, Austin: Cognitive Science Society.
- Wiley, J., Jarosz, A., Cushen, P., Jensen, M., & Griffin, T. D. (2009, November). The power of three: Why the third person matters. *Paper presented at the annual meeting of the psychonomic society*, Boston, MA.