Learning by Invention: Small Group Discussion Activities that Support Learning in Statistics

Andrew F. Jarosz, Olga Goldenberg & Jennifer Wiley

To cite this article: Andrew F. Jarosz, Olga Goldenberg & Jennifer Wiley (2016): Learning by Invention: Small Group Discussion Activities that Support Learning in Statistics, Discourse Processes, DOI: 10.1080/0163853X.2015.1129593

To link to this article: http://dx.doi.org/10.1080/0163853X.2015.1129593

Accepted author version posted online: 20 Jan 2016.
Published online: 20 Jan 2016.

Submit your article to this journal

Article views: 90

View related articles

View Crossmark data
ABSTRACT
Learning by invention is an alternative approach to teaching statistics where students are tasked with attempting to solve a problem before being taught the canonical formula for solving it, often resulting in increased understanding of material compared with traditional instruction. The first study, conducted in a college statistics classroom with mixed-skill groups, extended previous work showing an advantage for learning-by-invention activities compared with a lecture-first control. A second study explored group interactions that led to benefits from learning-by-invention activities. Successful groups were more likely to propose solutions and less likely to propose arbitrary formulas or highlight group members’ math skills. The features of small group interactions that may help or harm learning during invention activities are discussed.

Introduction
Teaching in mathematical courses, such as statistics, traditionally follows a standard routine. Students are introduced to a topic during a lecture, examples are worked through as a class, and then students practice the lesson and work through example problems on their own. Learning by invention is an alternative approach to teaching mathematical and statistical concepts by having students engage in small-group problem-solving activities where they are tasked with discovering their own methods of solving problems before they receive instruction on the canonical methods of solution. The goal of the present research is to examine how and when learning-by-invention activities may lead to better student understanding of statistical concepts and the role that group interaction may play in learning by invention by using discourse analytic methods to better understand discussion features that facilitate learning.

Invention, failure, and preparation for future learning
Recent interest in learning-by-invention paradigms can be traced to a popular study conducted by Schwartz and Martin (2004), comparing learning by invention against direct instruction during an instructional unit on statistics. Both instructional conditions received word problems that required comparing across two distributions and normalizing data. For students in the direct instruction condition, teachers introduced a procedure that could be used to solve the problems and then the students practiced this procedure on the given problems. However, students in the invention condition attempted to solve these problems in small groups without any instructional support. After the
problem-solving phase, students received a worked example of how to compute standardized scores. On a final transfer item requiring students to compute covariance, students in the invention condition significantly outperformed students who had been in the direct instruction condition. Schwartz and Martin interpreted these findings by suggesting that an invention activity can serve as preparation for future learning. They argued that engaging in invention activities helps students construct knowledge structures that in turn allows for better integration and uptake of the information available in the worked example. Also, taking various approaches as part of different attempts to solve the problem was suggested to help students attend to critical features of the problems.

Learning by invention has been also shown to be beneficial over direct instruction in several studies exploring the value of “productive failure” by Kapur (e.g., Kapur, 2012, 2014; Kapur & Bielaczyc, 2011). Kapur and colleagues have performed several studies examining the efficacy of having students try (and fail) to solve problems before learning the canonical solution. In these studies, students in the productive failure conditions often had less practice with problems than those in the direct instruction conditions, and yet students who engaged in problem solving before being taught a formula demonstrated significantly greater conceptual understanding and ability to transfer to novel problems than those who were taught the formula first (Kapur, 2012; Kapur & Bielaczyc, 2011).

Although the generation of a wide variety of solution attempts in the productive failure condition was found to predict learning in productive failure conditions in Kapur’s studies, it seems exposure to multiple solution attempts is not enough to account for all the benefits of invention activities. When a direct instruction condition incorporated examples of multiple solution approaches into the introduction of the canonical formula, this did not improve conceptual understanding to the same degree as the invention activities (Kapur, 2014). This suggests other features of the invention activity are contributing to its effectiveness. It is clear that learning by invention is beneficial under a variety of conditions, and, in particular, it benefits students in understanding material conceptually. However, the exact mechanisms through which it works are not yet fully understood. What remains to be demonstrated is why exactly learning-by-invention paradigms work and what steps may be taken to improve their impact.

**Exploring the small group context**

Some previous research has begun to explore the group contexts that may best facilitate learning by invention. Based on prior work suggesting that exposure to a wide range of solution approaches is part of what facilitates future learning, Wiedmann, Leach, Rummel, and Wiley (2012) tested the hypothesis that diversity among the members of a group may be particularly beneficial for reaping the benefits of learning-by-invention activities. Their results showed that groups consisting of both high and low math skill members (mixed groups) generated a broader range of solution attempts when asked to invent a formula for standard deviation than did more homogeneous math skill groups (groups with either all low or all high math skill members). Moreover, the wider range of solution alternatives helped students gain more from a subsequent lesson on the topic. Mediation analyses suggested that it was the discussion of a wider range of solution approaches during learning-by-invention activities, including the discussion of a larger number of higher quality solution attempts, that mediated the effects of group composition on learning. Notably, both high and low skill members seemed to benefit from participation in mixed groups as indexed by their score on a later quiz assessing their understanding of the standard deviation concept.

The results of Wiedmann et al. (2012) are consistent with many theoretical models of effective group collaboration. These propose that there is an advantage to working in diverse groups because it allows for possible synergy from multiple perspectives and a diverse, complementary knowledge base to draw on in problem solving (Brown, Tumeo, Larey, & Paulus, 1998; Canham, Wiley, & Mayer, 2012; Larson, 2010; van Knippenberg, De Dreu, & Homan, 2004). For example, the Brown et al. (1998) model predicts that groups consisting of members with different task-relevant knowledge should result in
cognitive stimulation from each other’s ideas due to the priming of ideas that otherwise have low accessibility.

However, despite theoretical predictions of synergy, a long tradition of research on the hidden profile paradigm (where each member is intentionally provided with unique information that is not shared by other members) shows that groups without additional instructions are usually unsuccessful in capitalizing on the diverse and unique cognitive resources potentially available to them. For example, it is well documented that groups will disproportionately favor the sampling and discussion of shared information compared with the unshared, unique information that can be critical for making an optimal decision or selecting an optimal solution approach (Stasser & Titus, 1985, 1987). Because of these negative forces in group discussions that may hinder the success of diverse groups, it is important to identify the conditions that can exacerbate and mitigate them.

**Overview of the studies**

Previous research on learning by invention has offered a great deal of evidence that this instructional method is promising and can be beneficial for student learning, but at the same time it leaves several open questions. In the present work the first question that is tested is whether previous findings of advantages in learning from invention activities in mixed groups can be replicated in the context of a college-level statistics course in comparison with a lecture-first control group. This extends the results of Wiedmann et al. (2012), which did not have a lecture-first condition. A second study examines which kinds of interactions among the members of mixed groups are most conducive for learning from these small group activities.

**Study 1**

Study 1 took place shortly after the first exam in an introductory statistics course in psychology. Students had learned about measures of central tendency and standard deviation but had not yet learned about standardization or $z$-scores. All students in this study were assigned to work in mixed groups consisting of both high- and low-skilled math individuals. In one condition the groups worked on an invention activity before hearing the lecture. In a second condition groups heard the lecture before completing the activity. The main prediction based on previous research was that students who participate in learning-by-invention activities before receiving a lecture should outperform those instructed by a more standard lecture-first approach.

One additional feature of small groups that might be important to consider is the role of the leader in the group process and outcome. Larson, Christensen, Franz, and Abbott (1998) demonstrated that leaders, more so than other members of the group, revisit the information previously contributed to the group discussion and ask questions about it. This can be an important predictor of learning. Leaders have also been shown to take charge of subtask assignments to better cope with task demands (Larson, 2010). Finally, leaders can offset the bias for shared information (e.g., Henningsen, Henningsen, Jakobsen, & Borton, 2004), which can play a critical role in exploration during group discussions.

Not all types of leaders, however, may be equally effective in ensuring the success of learning-by-invention activities. Expertise of the leader may be an important factor to consider. Past research shows that when groups are aware of who has particular expertise, the group is more likely to capitalize on their unique knowledge and skills (Franz & Larson, 2002). Wittenbaum, Hollingshead, and Botero (2004) found that experts are more likely to introduce valuable information into a group discussion when they are assigned a high status position, such as that of a group leader. Similarly, other research has shown that when the group leader holds unique information (i.e., the leader is the expert), more unshared information is uncovered and groups make better decisions (Cruz, Henningsen, & Smith, 1999; Henningsen et al., 2004). This research suggests that leader expertise may be an important predictor of effective information sharing in learning-by-invention tasks, which will subsequently affect learning outcomes.
Thus, in this study, groups were assigned either a high or low math skill member to serve as the leader. Based on prior research, a secondary prediction was that groups with high-skill math leaders should outperform those with low-skill math leaders.

**Methods**

**Participants**

Students enrolled in two different sections of an introductory statistics course in psychology at a large Midwestern university participated in a learning-by-invention activity on the general topic of variability as part of their normal coursework. Each section was taught by a different instructor. These instructors had roughly the same teaching experience, had previously collaborated in preparing lectures and exams, and often guest-lectured in each other’s courses, suggesting the two maintained a similar teaching style. Additionally, students from each class performed similarly on their first exam, \( t(88) = .37, \text{ ns} \). The larger section was assigned to the experimental manipulation (the learning-by-invention condition, \( n = 59 \)), whereas the smaller section served as a control (\( n = 31 \)). For topics related to this study, the two sections were taught using the same materials with students receiving identical instructions before the activity. Only students who completed all relevant activities were included in analyses. Because math ACT scores were unavailable for many of the students, students were categorized based on their scores from the first exam from the course and formed into mixed-skill groups (note that first exam scores correlated with math ACT scores for those students where it was available, \( r(59) = .32, p = .01 \)). Students scoring below the median were categorized as “low math skill,” whereas students above the median were categorized as “high math skill.”

**Materials**

As part of their normal Friday discussion section, students completed an invention activity on the general topic of variability and the specific topic of standardization and \( z \)-scores (based on the transfer item from Kapur [2012] and Wiedmann et al. [2012]; see Appendix A). The problem required students to determine the recipient of an award for the top science student at a school, with the caveat that the potential awardees were split across two different courses. They were then asked to identify who should win the award, determine a computation to make that decision, and to express that computation as a formula. Immediately before beginning the activity, students were given the following verbal instructions: “Today we’re doing a problem solving activity. I want you to use your statistics knowledge and see if you can answer the questions on this worksheet as a group. Do the best you can to answer these three questions, and write down all of the approaches that you try. I’ll collect the worksheets after 30 minutes. We will be going over this activity next Thursday in lecture.” Students in each section of statistics were already accustomed to group work during discussion sections, and the activity worksheets for this topic were similar to those normally encountered. Likewise, the lectures on standardization were the normal course lectures. Thus, nothing about the activity stood out to students as “different,” and the activity itself did not differ from what the instructors would normally have provided for the students.

**Learning outcomes.** As seen in Appendix B, students received a five-item quiz on standardization and \( z \)-scores. This quiz was intended to test students’ ability to both apply their knowledge of \( z \)-scores to statistical problems and demonstrate an understanding of the relationship between \( z \)-scores and raw scores. Four questions were multiple choice, covering simple calculations (e.g., calculate a person’s \( z \)-score based on a mean and standard deviation) and concepts (e.g., a \( z \)-score of zero is equivalent to a raw score equal to what?). The final quiz question followed a similar format to the activity worksheet students had completed during their discussion sections and required students to compare scores from individuals from different classes and decide which student did better relative to her class. Additionally, two items relating to these topics were on their next exam, several weeks later. One of these questions
required a comparison of students’ grades across classes with different means and standard deviations, whereas the other focused on the relationship between z-scores and raw scores.

**Procedure**

The timing of this worksheet was manipulated according to the section/instructor of the course: for one section (the “lecture-first” condition), students had already heard the normal lecture on the topic. The other section (the “learning-by-invention” condition) completed the worksheet during the discussion immediately before the lecture on standardization and z-scores. Thus, students in this condition had received lectures on all the material leading up to standardization but had not yet learned about standardization itself.

Students were assigned to groups composed of a mix of high- and low-skill students, with one student assigned as the group leader. Group leaders were selected with the goal of having approximately equal numbers of high- and low-skill leaders among the groups. The leaders were told that they needed to make sure that the group stayed on task and that everyone participated. Most groups contained three students, although due to absences some groups contained two or four students. There were a total of 32 groups, with 9 in the lecture-first condition and 22 in the learning-by-invention condition. Students were blind to all manipulations. Groups had 30 minutes to complete the activity.

In the learning-by-invention condition students completed the worksheet during the Friday discussion section the week before the class lecture on standardization and z-scores. The following week at the beginning of their lecture on standardization and z-scores, the instructor gave a brief explanation about the answers to the worksheet and discussed the solutions that students had come up with. After this, students received their normal lecture on standardization and z-scores. At the end of the lecture they received the quiz on the topic.

Students in the lecture-first condition first received their normal lecture on the topic of standardization and z-scores. During that week’s (post-lecture) Friday discussion section, they completed the worksheet. At the beginning of the next lecture (the first day of the next week of class), the instructor went over the answers to the worksheet and discussed the solutions that students had come up with. After this, students received their normal lecture on standardization and z-scores. The “solutions” introduced by the instructor during debriefing were the same in both conditions. After the worksheet debriefing, students completed the quiz.

**Coding**

**Learning outcomes.** Scores on the quizzes and performance on the relevant exam items were combined to create an overall learning score. Each quiz and test question was worth 2 points, allowing for 14 total points in students’ learning scores. Although most items were multiple choice, partial credit was possible on item 5 of the quiz, with one point awarded for having the proper calculations and one point for verbally stating the correct answer. These were then converted into a proportion of points out of 14 for analysis. Two raters independently graded these items with 100% agreement.

**Worksheet coding.** The worksheets from each group’s discussion were coded in an attempt to identify differences in the extent to which students engaged in explanation of the step-by-step processes they used while solving the problem. As seen in Table 1, six different steps necessary for the calculation of

<table>
<thead>
<tr>
<th>Step Number</th>
<th>Operation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Subtract the mean</td>
</tr>
<tr>
<td>2</td>
<td>Square the differences</td>
</tr>
<tr>
<td>3</td>
<td>Sum the squares</td>
</tr>
<tr>
<td>4</td>
<td>Divide by the number of data points</td>
</tr>
<tr>
<td>5</td>
<td>Take the square root</td>
</tr>
<tr>
<td>6</td>
<td>Calculate standard deviations from the mean</td>
</tr>
</tbody>
</table>
z-scores were identified. Two raters independently coded for the presence of each step, with 81% agreement. Disagreements were resolved by discussion between the two raters.

**Results**

Because the intraclass correlation for group members’ quiz scores were small (intraclass correlation = .11, \( p = .26 \), confidence interval = 95%), it was appropriate to analyze these data at an individual level. Table 2 provides learning outcomes data split by condition. A \( 2 \times 2 \times 2 \) (student math skill × leader math skill × activity condition) between-groups analysis of variance (ANOVA) was used to examine the impact of student math skill, leadership, and invention activities on learning scores.\(^1\) As seen in Figure 1, there were significant main effects of student math skill, \( F(1, 82) = 10.23, p = .002 \), partial \( \eta^2 = .11 \), and activity condition, \( F(1, 82) = 9.09, p = .003 \), partial \( \eta^2 = .10 \). High math skill individuals outperformed low math skill individuals, and those in the learning-by-invention condition outperformed those in the lecture-first condition. There was no significant effect of the leader’s math skill, \( F(1, 82) = 2.41, ns \), and no interactions between skill and math leadership, \( F(1, 82) = 2.65, ns \), or skill and activity condition, \( F(1, 82) = 1.65, ns \). There was a nonsignificant interaction between leader math skill and invention condition, \( F(1, 82) = 3.65, p = .06 \), partial \( \eta^2 = .04 \); however, this was subsumed by the three-way interaction, \( F(1, 82) = 4.20, p = .04 \), partial \( \eta^2 = .05 \). As seen in Figure 1, this interaction was driven by the pattern of results in the lecture-first condition, where low math skill individuals showed lower test performance when they worked in groups with high math skill leaders, \( t(12) = -2.31, p = .04, d = 1.33 \).

The worksheets from each group discussion were coded to identify differences in solution attempts. Almost all participants in the lecture-first condition used the information they learned in lecture on computing z-scores by calculating the standard deviations from the mean. However, few students in this condition wrote out the steps that led up to this (steps 1–5). On the other hand, students in the learning-by-invention condition were more likely to write out these earlier steps, and none actually arrived at the equation for z-scores. As shown in Figure 2, a \( 2 \times 2 \) (leader math skill × activity condition) ANOVA revealed a marginal effect for leader math skill, with groups with high math skill leaders showing fewer steps on their worksheets (\( M = 1.43, SD = .51 \)) than groups with low math skill leaders (\( M = 1.82, SD = .40 \)), \( F(1, 28) = 3.75, p = .06 \), partial \( \eta^2 = .12 \). There was no effect of activity condition and no significant interaction, \( F’s < 1 \).

**Discussion**

Replicating the results of several previous studies (Kapur, 2012, 2014; Kapur & Bielaczyk, 2011; Schwartz, Chase, Oppezzo, & Chin, 2011; Schwartz & Martin, 2004), students in the learning-by-invention condition outperformed students in the more traditional lecture-first condition on quiz and exam items related to the activity. A three-way interaction demonstrated that it was low-skill students

\(^{1}\)ANOVAs in Study 1 were calculated with Type III sums of squares to account for unbalanced groups.
in the lecture-first condition working in groups with high math skill leaders that led to this benefit in the learning-by-invention condition. In particular, although everyone did well in the learning-by-invention condition, in the lecture-first condition it was low math skill individuals in groups with high math skill leaders who showed a detriment on the quiz and exam. This suggests that low math skill individuals were the ones who seemed to benefit from participating in a learning-by-invention activity in mixed groups and that groups with high-skill leaders tended to be least likely to unpack the solution process as they engaged with worksheets.

As shown in the content analysis on the worksheets, groups with high math skill leaders tended to write down less information than those with low math skill leaders. This suggests that groups with high math skill leaders were more likely to skip steps while solving problems. Although the same lack of articulation was seen in the learning-by-invention condition, this skipping behavior may have been particularly harmful in the lecture-first condition and may be one reason why low-skill students did not benefit as much from small group discussion on practice problems after the lecture. One contribution of this study is showing that the benefits of invention activities can be found in an authentic course context, although at the same time the findings from this study are limited by the lack of true random assignment of students to conditions.

Figure 1. Learning score by activity condition, student math skill, and leader math skill for the standardization activity. Error bars represent one standard error of the mean.

Figure 2. Number of steps demonstrated on worksheets by leader math skill and activity condition. Errors bars represent one standard error of the mean.
Study 2

Study 1 replicated prior work that demonstrated benefits of engaging in learning-by-invention activities before receiving a lecture over a lecture-first-then-practice sequence. In contrast to Study 1, Study 2 was not an experimental study. Rather, the goal of Study 2 was to explore in more detail the discourse of groups during learning-by-invention activities. Discussion protocol data were analyzed to better understand how various features of small-group discussion during invention activities could affect learning outcomes.

One clear prediction from prior work is that we would expect more successful groups to generate and discuss more possible solution attempts (Kapur, 2014). By extension, we might expect a larger proportion of the discussion to be dedicated to discussing these proposals. A second prediction is that we might expect more successful groups to spend more time explaining concepts to each other. Webb (1980) found that high- and low-skill students working together can form mentor–mentee relationships. Such an arrangement can yield clear benefits for the low-skill students who are being mentored and would be unable to solve the problems on their own. In addition, however, advantages can also sometimes be seen for the high-skill students, as they benefit from teaching others. The work of Webb and others has suggested that group effectiveness may depend largely on the amount of time that students spend explaining math concepts to each other (e.g. Webb, 1980). In addition, one might expect that more successful groups might spend more time in task coordination (Moreland & Levine, 1992) or evaluating other’s proposals (Barron, 2003; Schoenfeld, 1989) and less time in off-task talk (Dugosh, Paulus, Roland & Yang, 2000; Wiley & Jolly, 2003). Prior research in the group literature suggests that groups who recognize the expertise of a group member should have more useful discussions (Franz & Larson, 2002). This would suggest that more successful groups may be more likely to attempt to identify who might have relevant math expertise among members.

Following the theoretical implications of the productive failure approach (Kapur, 2012), one might also predict that groups who explicitly recognize they are at impasse or remark on experiencing failure might be more likely to be receptive to the canonical solution when it is given to them after the activity. This prediction would also be consistent with prior work that has shown that having experienced some form of impasse leaves cues related to the major issues of the problem in memory, leaving the solvers more receptive to the solution when it is eventually discovered or presented (Gick & McGarry, 1992; Seifert, Meyer, Davidson, Paralano, & Yaniv, 1995).

Methods

Participants

Participants were 14 groups of three individuals enrolled in introduction to psychology at a large Midwestern university. These students were recruited via a psychology research participant pool outside of normal class time and completed the study for partial course credit. Each group was heterogeneous in terms of math skill, based on students’ math ACT scores, with students scoring higher than a 25 on their math ACT being categorized as “high math skill.” Students were specifically recruited to ensure that each group had at least one high math skill individual and one low math skill individual. Of the 14 groups, the balance of high and low math skill individuals was split equally, with seven groups containing two low math skill individuals and seven groups containing one low math skill individual.

Materials

Learning by invention. Because standardization requires knowledge of standard deviation and introductory psychology students are unlikely to already know this formula, students completed an invention activity on the general topic of variability and on the specific topic of standard deviation. Standardization served as a transfer task after this activity. The worksheet (based on Kapur [2012] and described in Wiedmann et al. [2012]) is presented in Appendix C. The problem involved trying to find out which tea-growing company produced tea with the most consistent year-to-year levels of
antioxidants. Students were asked to either write down a formula to calculate this consistency or to write step-by-step instructions on calculating consistency (some students are overwhelmed by being asked for a formula; Wiedmann et al., 2012).

**Learning outcomes.** Learning was determined using a quiz containing three open-ended items (Wiedmann et al., 2012), as seen in Appendix D. Two of these quiz items required applying the concept of standard deviation to determine consistency in weather, whereas one quiz item was a transfer item (asking students to derive standard scores using the problem used on worksheets in Study 1 and found in Appendix A).

Quizzes were scored according to the same criteria as used by Wiedmann et al. (2012). Points were assigned according to the concepts referenced within students’ quiz answers for each item, as listed in Table 3. Each item received the point value of the highest response code applicable to their response. In total, the quiz was worth up to 12 points. Each item of the quiz was scored by two different graders to allow for an estimate of reliability. Discrepancies were resolved by a third rater. Reliability was good for each item, with Krippendorff’s $\alpha$ of .84, .76, and .69 for the first, second, and third quiz items, respectively.

**Procedure**

In contrast to Study 1 all participants received the learning-by-invention manipulation in Study 2. After this they completed the invention activity about tea growers. They were given 30 minutes to complete this activity. Students were then separated to work individually and were given an overview of the concept of standard deviation as well as a worked example of a standard deviation problem. Finally, they completed the quiz. This entire process took approximately 1.5 hours to complete.

**Discourse coding**

Because the goal of this study was to understand how the features of group discourse impacted learning, utterances made by each group member were coded into one of the following mutually exclusive categories: (1) solution proposals; (2) clarification request or response; (3) evaluative comments; (4) comments related to group task coordination; (5) calculations; (6) comments on math skill or expertise; (7) comments about being stuck, at impasse, or at a lack of forward progress; or (8) off-task comments. Krippendorff’s $\alpha$ indicated good interrater reliability on all coding metrics (> .77).

**Results**

An overview of the frequency of groups’ utterances per coding category can be seen in Table 4. Only a small proportion of these utterances were solution proposals. Almost half of the comments were clarifications or requests for clarifications about a proposed solution. Roughly 16% of utterances were evaluations of suggested approaches, whereas off-tasks comments made up 11% of utterances.

Because the intraclass correlation between group members’ quiz scores was small (intraclass correlation $= .08$, $p = .30$, confidence interval $= 95\%$), it was appropriate to analyze these data at an individual level. Despite all groups participating in the learning-by-invention condition, in this sample there remained a gap between low ($M = .67$, $SD = .17$) and high math skill ($M = .83$, $SD = .13$) students’ performance on the learning outcomes task, $t(40) = 3.40$, $p = .002$, $d = 1.05$. However, some

<table>
<thead>
<tr>
<th>Table 3. Quiz Response Codes.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Response Code</strong></td>
</tr>
<tr>
<td>Single point estimate</td>
</tr>
<tr>
<td>Ranges and deviations</td>
</tr>
<tr>
<td>Incorrect SD</td>
</tr>
<tr>
<td>Correct SD</td>
</tr>
</tbody>
</table>
low math skill individuals performed as well as high math skill individuals, whereas others performed more poorly. To further investigate possible reasons for these divergent outcomes, groups were divided into “more successful” and “less successful” categories according to how well the low math skill members in those groups did on the learning outcome measure. For each group, the low math skill group members’ performance was averaged. After this, a median split was used to divide the groups. To ensure this resulted in an appropriate categorization of the different groups, a $2 \times 2$ (more vs. less group success, high vs. low math skill) between-subjects ANOVA was conducted at the individual participant level. As seen in Figure 3, there were main effects of math skill, $F(1, 38) = 17.99, p < .001$, partial $\eta^2 = .32$, and group success, $F(1, 38) = 8.24, p = .007$, partial $\eta^2 = .18$. However, these effects were driven by a significant math skill by group success interaction, $F(1, 38) = 12.23, p = .001$, partial $\eta^2 = .24$. In more successful groups, low- and high-skill group members did not differ in performance, $t(19) = .51$, ns, whereas in less successful groups, low math skill individuals performed significantly worse, $t(19) = 5.71, p < .001, d = 2.51$. Finally, it should be noted that there was little difference in group composition between more and less successful groups. Of the more successful groups, three of those groups contained two high math skill members, whereas the less successful groups included four groups with two high math skill members. This suggests that group composition itself was not driving these differences in success.

To explore differences between more successful and less successful groups, several aspects of their discourse were considered. First, discourse was coded in terms of the total number of utterances made by each individual. Individuals in more successful groups made fewer utterances ($M = 39.57$,
SD = 24.66) than those in less successful groups (M = 75.76, SD = 45.04), t(40) = 3.23, p = .002, d = 1.02. Thus, the remaining analyses were conducted in terms of the proportion of an individual’s utterances falling into each mutually exclusive coding category, as seen in Table 5.

After ensuring there were no issues of multicollinearity (with the highest correlation between proportion of utterances being between off-task and clarification comments at \( r = -.62 \)), the proportions for each category were submitted to a MANOVA. This resulted in a significant effect of group success, \( F(8, 33) = 3.33, p = .007 \), partial \( \eta^2 = .45 \). As seen in Table 5, for both types of groups most utterances fell into clarification, evaluation, and calculation categories. Most comments consisted of asking a person to repeat the prior utterance or saying numbers or operations out loud. Using independent samples t-tests for follow up comparisons, there were no significant differences in individuals’ proportion of utterances that fell into the clarification, evaluation, or calculation categories for more or less successful groups. Comments referring to impasse or failure were rare and did not vary by group success.

There were significant differences in solution proposals, comments related to coordination, math expertise, and off-task talk. The more successful groups spent a greater proportion of their utterances discussing solution proposals compared with less successful groups. On the other hand, more successful groups had a lower proportion of their statements related to coordination, a lower proportion of utterances related to math expertise or skill, and a lower proportion of off-task utterances.

**Does it matter who made the utterances?**

One interesting question is whether it matters who contributed each of these utterance types (high- vs. low-skill members). To examine this, the effects of math skill and its interaction with group type were explored in ANOVAs for each discourse category. In no case were there any effects seen for math skill or any interactions. This suggests that the source of each of these utterance types did not affect the success of the group.

**Discussion of previously known mathematical concepts**

Although the students in this sample had not yet been taught about standard deviation as part of the introduction to psychology course, there were four groups in which “standard deviation” was brought up during the discussion as a potential solution. In the initial coding these were coded as solution proposals; however, because standard deviation is in fact the solution method for this problem, these instances warranted closer inspection. Simply mentioning standard deviation was not enough to ensure the groups were successful in helping the lowest math skill students to learn the concept. Of the four groups that mentioned standard deviation, only two were successful. In one successful group the high math skill member both introduced the standard deviation formula and explained it to others in the group. In the other successful group, there were two low math skill members who each believed that standard deviation seemed relevant, but neither could remember the equation.

### Table 5. Proportion of Utterances in Each Coding Category by Group Success.

<table>
<thead>
<tr>
<th>Proportion of Group Utterances</th>
<th>Less Successful</th>
<th>More Successful</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Solution proposals</td>
<td>.05</td>
<td>.04</td>
</tr>
<tr>
<td>Clarifications</td>
<td>.46</td>
<td>.11</td>
</tr>
<tr>
<td>Evaluations</td>
<td>.16</td>
<td>.06</td>
</tr>
<tr>
<td>Coordination</td>
<td>.03</td>
<td>.04</td>
</tr>
<tr>
<td>Calculations</td>
<td>.10</td>
<td>.06</td>
</tr>
<tr>
<td>Math skill</td>
<td>.01</td>
<td>.02</td>
</tr>
<tr>
<td>Impasse</td>
<td>.01</td>
<td>.02</td>
</tr>
<tr>
<td>Off-task</td>
<td>.08</td>
<td>.11</td>
</tr>
</tbody>
</table>

DISCOURSE PROCESSES
Instead, they worked together to come up with an approach that approximated standard deviation by using average deviations from the mean. In one unsuccessful group there were two high math skill members who thought that the answer was to compute the standard deviation, but neither could remember the formula. However, they did not pursue the idea further. Finally, in the last unsuccessful group a high math skill member knew the standard deviation formula but offered no explanation of it to the others.

Although the proposal or discussion of using standard deviation to determine consistency did not differentiate the more successful from the less successful groups, the discussion of arbitrary formulas did. Many groups discussed math concepts that were not particularly appropriate for determining the consistency of the data including using slope, derivatives, area under the curve, the quadratic formula, and linear regression. To capture this behavior, each discussion was coded for the number of utterances in which members said the words formula, equation, or prove, including utterances in which an arbitrary formula name or equation was generated (e.g., “\(y = mx + b\)”). This analysis revealed that the members in less successful groups were more likely to speak in vague terms about formulas or to offer arbitrary formulas into the discussion (\(M = 4.19, SD = 3.64\)) than members in successful groups (\(M = 1.57, SD = 2.18\)), \(t(40) = 2.83, p = .007, d = .89\).

**Discussion**

Consistent with prior work showing that discussing a larger number of solution alternatives facilitates learning from invention activities, successful groups spent a larger proportion of their time discussing possible solutions. At the same time more successful groups made fewer overall utterances than less successful groups, as the less successful groups spent more time off-task. The less successful groups also spent more time on coordinating the task and discussing math skill and arbitrary formulas. These last two findings were not expected based in prior work and are discussed in more detail below. One aspect of this study worth noting is that it did not matter who made the comments of various types. Another is that explanation did not seem to occur between group members as they worked in this activity. Finally, reaching an impasse during discussion seemed to have no impact on reaping benefits from the task. Taken together, these results suggest that groups who were most likely to benefit from the learning-by-invention task were those that focused on discussing solution proposals and who avoided task-irrelevant discussion.

**General discussion**

Two main findings come from these studies. First, Study 1 showed that students who participated in learning-by-invention activities in mixed-skill groups in an authentic course context outperformed those who received a traditional lecture-practice sequence on later tests. Second, Study 2 demonstrated that low math skill members who were in groups where a larger proportion of the group discussion was spent evaluating possible solutions learned more from the activity. Two specific features of discussion that seemed to undermine the effectiveness of learning-by-invention activities were the preoccupation of group members with math skill and with the goal of generating a formula.

These observations suggest that it may be important to provide students with some direction during their discussions to ensure students are clear about the goal of the invention activity. It also seems important to encourage students to share their understanding by walking step-by-step through calculations in their group work. For example, a student doing calculations on her own and then giving a partial result without showing the work that went into calculating it is unlikely to provide a benefit to the other members of the group. The low math skill individuals seemed to suffer when their groups failed to engage in this kind of interaction in Study 1. By pushing students to walk through their calculations in more detail while they are working toward a solution, group members who otherwise would not have known those steps are able to benefit. Even if the calculations used are inappropriate for the actual solution, this would serve to highlight the difficulties of the problem at hand for the group as
a whole, which may prove beneficial when learning the canonical solution later. Note that this is not the same as giving them a worked example; students are still tasked with discovering the solution on their own. It does, however, make thinking explicit and share both successes and failures in solutions across the group as a whole.

In Study 2 students did not seem to engage in explanation at all. The fact that discussion of math skill was linked to less successful learning raises the possibility that students may have experienced some form of math anxiety or stereotype threat (a fear of confirming a negative stereotype) while participating in this task. If a student is already unsure of his math abilities, then making those abilities salient through direct discussion may lead to increased anxiousness or to ruminations about math ability. Combined with the fact that being asked to discover a new mathematical approach may seem especially daunting to some students, this shift in focus could reduce students’ ability to benefit from a learning-by-invention paradigm. Additional instructions or interventions may need to be added to invention activities to help them seem less threatening (Maloney, Schaeffer, & Beilock, 2013).

Several insights from these results resonate with other recent approaches to learning-by-invention activities that have highlighted the importance of providing some scripting or structure to the group discussions to help students to get the most from these learning opportunities (Kapur & Bielaczyc, 2011; Roll, Holmes, Day, & Bonn, 2012; Westermann & Rummel, 2012). For example, it has been argued that scripting the roles of students, such as having a “thinker” and a “questioner,” in learning-by-invention activities may increase what those students can get out of the activity (Westermann & Rummel, 2012). Similarly, metacognitive scaffolding provides an environment that encourages students to engage in more exploratory analyses, self-explanations, and peer interactions, and this has been shown to increase the number of conceptual features students address as well as the number of times that students revise their approach (Roll et al., 2012). Moving forward, future studies are still needed to establish the most effective ways to implement learning-by-invention activities and to explore additional manipulations in a classroom setting that can push group dialogue toward productive interactions.

In theoretical terms, invention activities have been discussed both as providing preparation for future learning (Schwartz & Martin, 2004) and as opportunities for productive failure (Kapur, 2012). Although some work shows that experiencing failure or reaching impasse can make individuals more receptive to new information (Gick & McGarry, 1992; Seifert et al., 1995), Study 2 was unable to provide support for the idea that failure might be critical for successful performance on learning-by-invention activities, at least in terms of evidence from spontaneous comments of groups being at impasse or in a state of failure. However, as the amount of discussion dedicated to solution alternatives did relate to group success, the results of these studies offer additional support for the idea that considering a range of possible solutions during discussion lays the groundwork for later uptake of the canonical formula and facilitates learning from these activities.

From an instructor’s standpoint, these activities are relatively simple to include in the curriculum, and in many cases they may be quite similar to those already being used in the classroom, with the only major difference being the timing of the direct instruction phase versus the activity phase. Learning-by-invention activities also appear to be highly motivating (Belenky & Nokes, 2013), which also bears further research. In an early recommendation for this style of learning activity, Wertheimer (1959) discussed how children who have been taught how to compute the area of a parallelogram will blindly use the formula and fail to grasp what they have been taught. He contrasted this with children who discover how to transform the slanted figure into a rectangle, via the insight to “cut off” one of the pointed ends, and to use it to complete the other pointed end, so they can then see the application of the usual formula for the area of a rectangle. These students learned more than just a rote procedure. Rather, they truly grasped the mathematical relations and were able to apply and use the formula with understanding. Although the promises of learning from discovery and invention activities such as these have been discussed for quite a long time, more work remains to be done to understand when they may actually prove to be most successful.
References


Appendix A

Study 1 Invention activity (adapted from Kapur, 2012)

Two senior students were nominated for the “Best Science Student” award for 2009. Kelvin White is the top physics student, whereas Alicia Kwan is the top chemistry student for 2009. Table A1 shows the physics and chemistry top scorers between 2004 and 2009, with their scores presented in ascending order.

Table A1.

<table>
<thead>
<tr>
<th>Name</th>
<th>Year</th>
<th>Score</th>
<th>Name</th>
<th>Year</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tham Ling</td>
<td>2004</td>
<td>82</td>
<td>Abdul Basher</td>
<td>2005</td>
<td>82</td>
</tr>
<tr>
<td>Jodie Hampton</td>
<td>2007</td>
<td>83</td>
<td>Fredrick Chay</td>
<td>2004</td>
<td>85</td>
</tr>
<tr>
<td>Jeremy Butler</td>
<td>2003</td>
<td>83</td>
<td>Linda Powell</td>
<td>2006</td>
<td>88</td>
</tr>
<tr>
<td>Chee Foster</td>
<td>2006</td>
<td>84</td>
<td>Terry Watson</td>
<td>2008</td>
<td>91</td>
</tr>
<tr>
<td>Susan Teo</td>
<td>2005</td>
<td>84</td>
<td>Noah Osai</td>
<td>2007</td>
<td>95</td>
</tr>
<tr>
<td><strong>Kelvin White</strong></td>
<td>2009</td>
<td><strong>94</strong></td>
<td><strong>Alicia Kwan</strong></td>
<td>2009</td>
<td><strong>99</strong></td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>85</td>
<td>Mean</td>
<td></td>
<td>90</td>
</tr>
</tbody>
</table>

Both Kelvin and Alicia are the best performers in their respective subjects for the past 6 years.

1. Who do you believe should win the award? Use your statistics skills to back up your answer!
2. Come up with a computation to help you make this decision.
3. Write your computation as an equation or formula that could be applied to other data sets.

Appendix B

Study 1 Quiz and exam items

Quiz items

1. Jenna scored 40 on a standardized test of reading ability where the mean score is 50 and the standard deviation is 10. Based on this information, what is Jenna’s z-score?
   A. − 2.0
   B. 2.0
   C. − 1.0
   D. 1.0

2. A person with a z-score of zero would have a raw score equal to
   A. the lowest score in the distribution of raw scores
   B. the mean of the distribution of raw scores
   C. the highest score in the distribution of raw scores
   D. zero

3. Following is a list of z-scores from a single distribution of scores. Which of the z-scores corresponds to the raw score farthest from the mean of the distribution?
   A. − 2.3
   B. − 1.5
   C. 0.8
   D. 1.2

4. Daniel wanted to know his approximate score on the final exam for his mathematics class. His professor hinted that his score was well above the class average. The professor announced that the mean for the class final exam was 90 with a standard deviation of 7. Given Daniel’s z-score of 1.67, what is the raw score for Daniel’s exam grade?
A. 100.00  
B. 101.69  
C. 102.45  
D. 88.17

5. Two students recently took algebra class tests. The students are at different schools but wanted to compare their performance. The first student scored 80 on the test. Her class average was 90 with a standard deviation of 10. The second student scored 70. Her class average was 50 with a standard deviation of 10. Which student did better? Show your work and circle your final answer.

Exam items

1. Dan got a 60 on his social psych exam. The class average was 70 with $SD = 5$. Chris got an 80 on his cognitive psych exam. The class average was 60 with $SD = 10$. Who did better, compared with the rest of his class?
   A. Dan  
   B. Chris  
   C. Both students did equally well  
   D. There is not enough information to answer this question

2. A person with a z-score of zero would have a raw score equal to
   A. the lowest score in the distribution of raw scores  
   B. the highest score in the distribution of raw scores  
   C. zero  
   D. the mean of the distribution of raw scores

Appendix C

Study 2 Invention activity (from Wiedmann et al., 2012)

Mr. Fergusson, Mr. Merino, and Mr. Eriksson are the managers of the Tea Company. They are searching for a new supplier of green tea, and after a long search they shortlisted three potential growers in India: Thourbo, Dareen, and Ging. All growers are asking for the same price for their tea. Because the Tea Company is advertising the healthy properties of their product, the managers agreed that they should base their decisions on the level of antioxidants found in each tea for the last six harvests. Table C1 shows the amount of antioxidants that the tea from each grower contained between 2002 and 2007. However, Thourbo didn’t get the equipment for testing antioxidant levels until 2003, so the 2002 level is unknown.

<table>
<thead>
<tr>
<th>Year</th>
<th>Thourbo (antioxidants per mg)</th>
<th>Dareen</th>
<th>Ging</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td>14</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td>10</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>2004</td>
<td>13</td>
<td>14</td>
<td>17</td>
</tr>
<tr>
<td>2005</td>
<td>17</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>2006</td>
<td>20</td>
<td>19</td>
<td>14</td>
</tr>
<tr>
<td>2007</td>
<td>15</td>
<td>18</td>
<td>21</td>
</tr>
</tbody>
</table>

The managers agreed they should buy the tea with the most consistent levels of antioxidants from year to year. They decided to approach this decision mathematically and want an equation or formula for calculating the consistency of antioxidant levels for each tea grower. This formula should apply to
each of the tea growers and help to provide a fair comparison between them. The managers decided to get your help.

Your task is to come up with a formula for consistency that will allow you to show which tea grower has the most consistent levels of antioxidants. Your group can come up with more than one formula or equation, but in the end circle the one you decide on.

Appendix D

Study 2 Quiz (from Wiedmann et al., 2012)

\[
\text{Standard deviation} = \sqrt{\frac{\sum (x - M)^2}{n}}
\]

In preparing for an ice hockey tournament in 2009, the organizers had to decide which month in which to hold the event. With an outdoor ice rink, the less variability in temperature from day to day, the less it costs to maintain the ice. The organizers therefore wanted to choose a month with the most consistent temperatures for this event. They narrowed their options to January and February and decided to examine daily temperature for six randomly selected days in each month in 2008 to make their choice. The high temperatures on each of those days (in Fahrenheit) for the 2 months are shown below in Table D1.

<table>
<thead>
<tr>
<th>Week 1, Day 1</th>
<th>January (°F)</th>
<th>February (°F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 2, Day 2</td>
<td>32</td>
<td>31</td>
</tr>
<tr>
<td>Week 2, Day 7</td>
<td>35</td>
<td>34</td>
</tr>
<tr>
<td>Week 3, Day 7</td>
<td>33</td>
<td>37</td>
</tr>
<tr>
<td>Week 4, Day 5</td>
<td>41</td>
<td>33</td>
</tr>
<tr>
<td>Week 4, Day 6</td>
<td>39</td>
<td>44</td>
</tr>
</tbody>
</table>

Based on the information in the table, which month should the organizers choose, given that they would want a month that has the most consistent temperatures? Explain your decision mathematically.

A few days later the organizers relooked at the data and realized they made a mistake for the figure recorded Week 4, Day 5 in January. Instead of 41°F (see the table below), the maximum temperature should be 60°F. Given this new figure, which month should they choose now if they want one that has the most consistent temperatures? Why did this mistake matter (or not)?

<table>
<thead>
<tr>
<th>Week 1, Day 1</th>
<th>January (°F)</th>
<th>February (°F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 2, Day 2</td>
<td>32</td>
<td>31</td>
</tr>
<tr>
<td>Week 2, Day 7</td>
<td>35</td>
<td>34</td>
</tr>
<tr>
<td>Week 3, Day 7</td>
<td>33</td>
<td>37</td>
</tr>
<tr>
<td>Week 4, Day 5</td>
<td>41</td>
<td>60</td>
</tr>
<tr>
<td>Week 4, Day 6</td>
<td>39</td>
<td>44</td>
</tr>
</tbody>
</table>

Two senior students were nominated for the “Best Science Student” award for 2009. Kelvin White is the top physics student, whereas Alicia Kwan is the top chemistry student for 2009. Table D2 shows the physics and chemistry top scorers between 2004 and 2009, with their scores presented in ascending order.
Both Kelvin and Alicia are the best performers in their respective subjects for the past 6 years. Because there is only one “Best Science Student” award, who do you believe deserves the award more? Please explain your decision mathematically.

Table D2.

<table>
<thead>
<tr>
<th>Top Physics Students</th>
<th>Top Chemistry Students</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Name</strong></td>
<td><strong>Year</strong></td>
</tr>
<tr>
<td>Tham Ling</td>
<td>2004</td>
</tr>
<tr>
<td>Jodie Hampton</td>
<td>2007</td>
</tr>
<tr>
<td>Jeremy Butler</td>
<td>2003</td>
</tr>
<tr>
<td>Chee Foster</td>
<td>2006</td>
</tr>
<tr>
<td>Susan Teo</td>
<td>2005</td>
</tr>
<tr>
<td><strong>Kelvin White</strong></td>
<td><strong>2009</strong></td>
</tr>
<tr>
<td>Mean</td>
<td>85</td>
</tr>
</tbody>
</table>