Outcomes and Mechanisms of Transfer in Invention Activities

Ido Roll (ido@phas.ubc.ca)
Carl Wieman Science Education Initiative, University of British Columbia
6224 Agricultural Road, Vancouver, BC V6T-1R9, Canada

Vincent Aleven (aleven@cs.cmu.edu)
Human Computer Interaction Institute, Carnegie Mellon University,
5000 Forbes Avenue, Pittsburgh, PA 15213, USA

Kenneth R. Koedinger (koedinger@cmu.edu)
Human Computer Interaction Institute and Department of Psychology, Carnegie Mellon University,
5000 Forbes Avenue, Pittsburgh, PA 15213, USA

Abstract

Invention activities are structured tasks in which students create mathematical methods that attempt to capture deep properties of data (e.g., variability), prior to receiving instruction on canonical methods (e.g., mean deviation). While experiments have demonstrated the learning benefits of invention activities, the mechanisms of transfer remain unknown. We address this question by evaluating the role of design in invention activities, identifying what knowledge is acquired during invention activities, and how it is applied in transfer tasks. A classroom experiment with 92 students compared the full invention process to one in which students evaluate predesigned methods. Results show that students in the full invention condition acquired more adaptive knowledge, yet not necessarily better procedural knowledge or invention skills. We suggest a mechanism that explains what knowledge invention attempts produce, how that knowledge is productively modified in subsequent instruction, and how it improves performance on some measures of transfer but not others.

Keywords: Invention activities, transfer, intelligent tutoring systems, modular knowledge, generation.

Introduction

Invention activities ask students to design and evaluate mathematical methods that capture deep properties of given examples. For instance, the task in Figure 1 asks students to invent a general method for calculating variability. Invention activities are designed to augment and precede traditional teacher-led instruction (Roll, Aleven, & Koedinger, 2009; Schwartz & Martin, 2004). Following the invention attempt, whether successful or not, students receive instruction on canonical methods for the same problem ("show"), and apply these methods to different problems ("practice"). For example, after inventing measures of variability, students receive show-and-practice instruction on Mean Deviation (MD), that is, the mean absolute difference from the mean. One key aspect of invention activities is the use of contrasting cases (Chase, Shemwell, & Schwartz, 2010). Contrasting cases are carefully designed examples that emphasize target features by changing just those and no other features. For example, the contrasting cases shown in Figure 1 (middle) emphasize distribution while fixing sample size, range, average, etc.

The Invention Lab facilitates invention activities in three steps (Roll, Aleven, & Koedinger, 2010). Students are first asked to rank the contrasting cases according to the target property (e.g, the variability of the left graph is lower than that of the right graph, see Figure 1). Students then design a mathematical method and calculate the target property for the given data (e.g., design "range / N" and apply it to both data sets). Last, students evaluate their method by comparing its inferred ranking to the initial qualitative ranking. Once the invented method ranks the contrasting cases successfully, students are given new data to work with. Students often make progress during the invention process, yet they rarely invent a valid general method (Schwartz & Martin, 2004).

Classroom evaluations found that the combination of invention activities and show-and-practice instruction improves performance on transfer measures, compared with show-and-practice alone, controlling for overall time on task (Schwartz & Martin 2004; Roll et al., 2009; Kapur, 2008). However, while the positive effect of invention activities is well documented, not enough is known about how this effect is achieved. In this paper we investigate the cognitive processes and knowledge outcomes that are associated with invention activities, and the manner in which invention activities foster performance on measures of transfer.

Invention vs. Evaluation

Our first goal is to evaluate the role of a key process within the invention activity – the design of methods. Hypothesis 1(a) suggests that the design of methods is a key process in achieving the results of invention activities, akin to the generation effect (Richland et al., 2005). However, the worked example effect suggests that novice learners may benefit from analyzing given solutions more than from engaging in problem solving (Sweller, 2006). In addition, evaluating given methods may take less time than designing and evaluating methods. Therefore, Hypothesis 1(b) suggests that evaluation of given (imperfect) methods, using contrasting cases, is more beneficial than designing methods.

Knowledge Outcomes and Mechanisms of Transfer

Our second goal is to identify the knowledge that is acquired during invention activities and enables transfer. This is especially interesting given that students usually fail to generate successful methods. Hypothesis 2(a) suggests that students acquire domain-specific knowledge that prepares them to better encode the subsequent instruction (Schwartz & Bransford, 1998; Schwartz, Sears, & Chang, 2007). Hypothesis 2(b), on the other hand, suggests that students acquire domain-independent habits or skills that help them approach transfer items. For example, Taylor et al. (2010) found that invention activities help students come up with more ideas during open-ended transfer problems (yet the quality of each idea was not affected by the invention process).

Last, we propose a mechanism that explains how knowledge that is acquired during invention activities interacts with show-and-practice instruction to improve ability to transfer.

Method

Design

We address these questions by comparing two versions of invention activities. Students in the *Design and Evaluate* condition were instructed to design methods and evaluate them, while students

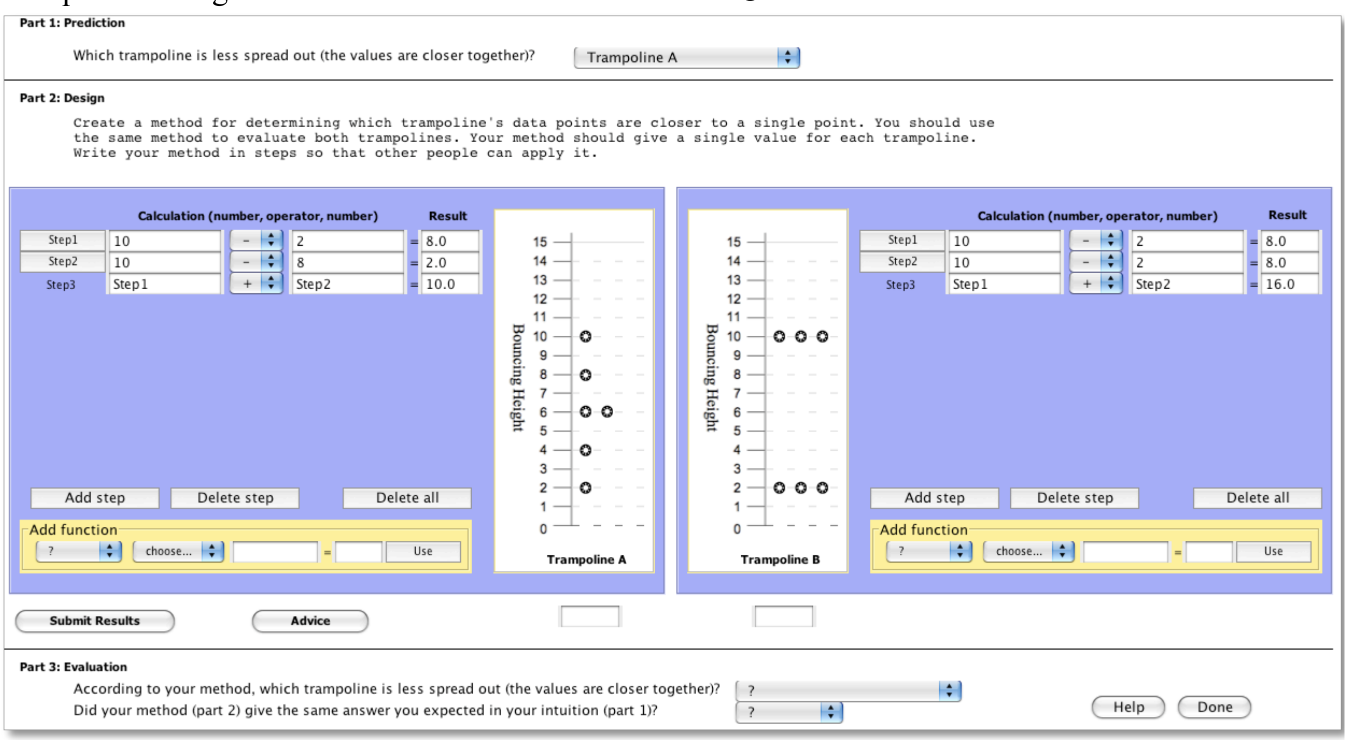

Figure 1: The Invention Lab. Students invent methods for calculating variability by qualitatively ranking the contrasting cases, inventing a mathematical procedure, and comparing their initial ranking with their invented procedure. The cover story compares the consistency of two trampolines. The contrasting cases target students' demonstrated knowledge gaps.

in the *Evaluate Only* condition were asked evaluate predesigned methods. These predesigned methods were taken from a previous study in the same school with the same age group of students (Roll, et al., 2009). Each of the chosen methods failed to incorporate one or more critical features of the target domain, variability. For example, one method that was given to students was "range / N". This method fails to use all data points in each of the contrasting cases, and thus fails to generate a correct ranking for the cases shown in Figure 1. Students were instructed to test each method against multiple sets of data. As soon as a method was found not to work, students were asked to move to the next method.[1]

Participants.

Ninety-two 7th grade students from a public middle school participated in the study. 33 students were randomly assigned to the *Design and Evaluate* condition and 59 students were assigned to the *Evaluate Only* condition. Students were enrolled in two levels of math classes (regular and advanced), split evenly between conditions.

Procedure

The study spanned four class periods over two days. The first day began with a pre-test, followed by an Invention Lab tutorial. Students then moved on to the first invention activity, which was limited to 25 minutes. Instruction for the invention activity differed based on condition. *Design and Evaluate* students were asked to design and evaluate methods, while *Evaluate Only* students were asked to implement and evaluate methods from a given booklet. Students invented in pairs and chose their partners. With few exceptions, the pairs did not change throughout the study. The invention activity concluded with a short whole-class discussion in which students (from both conditions) shared the successes and limitations of their methods. The discussion was used to motivate students to try their best during the invention process. The second invention activity followed a similar structure and duration, and was divided across two days. Following the summary discussion of the second activity, students received about 7 minutes of instruction on MD. Students then worked individually with a designated intelligent tutoring system for MD. All students finished all practice problems within 25 minutes. The study concluded with a post-test.

Materials

Students in both conditions used the Invention Lab (Roll et al., 2010), an intelligent tutoring system created using the Cognitive Tutor Authoring Tools (Aleven et al, 2009). The first invention activity asked students to evaluate which of two trampolines is more consistent, given data from multiple bounces for each trampoline (see Figure 1). The first set of contrasting cases emphasized the range of the data

Procedural fluency:
What is the MD of 2, 4, 7, and 3? *(1)*

Conceptual knowledge:
The average temperature in Montreal and Vancouver is similar, but the MD is much higher in Montreal. What does it mean?
a. Montreal is always warmer.
b. Vancouver is always warmer.
(c). The temperatures in Montreal are more extreme.
d. The temperatures in Vancouver are more extreme.

Debugging items:
Marlene invented the following method:
Step 1: Find the average.
Step 2: Find the distances from the average.
Step 3: Add up all the distances.
What do you think about this method?
a. It works.
(b). It does not work because there is different number of numbers in each set.
c. It does not work because the method does not use all the numbers.

PFL items:
Sonly is testing a new distance detector. The device is not accurate and small errors are permitted. Sonly measured the same distance several times and received 3, 5, 6, and 10 feet. What is the MD of the new device, if errors of 1 foot or less should be ignored? *(1.75)*

Figure 2: MD assessment items. Correct answers appear in parentheses. All items but procedural fluency items were given in the context of story problems.

[1] Half of the *Evaluate Only* students were also prompted to explain why each method failed or succeeded. We collapse these groups because it was evident that students ignored these prompts and we found no other associated differences in learning.

(1,3,5,7,9 vs. 3,4,5,6,7). Subsequent contrasting cases were created by the Invention Lab in real time based on the weaknesses of students' methods. The second invention activity asked students to evaluate the consistency of machines that pack candies into bags in a candy factory.

Instruction on how to apply MD was delivered by the first author and included graphical and mathematical explanations of the structure of the formula and a few examples. In subsequent practice students applied MD to 20 different data sets and interpreted results.

Measures and Analysis

The pre- and post-tests included assessments of procedural and conceptual knowledge. The post-test also included two types of transfer items that required students to construct or analyze modified versions of the taught procedure: Preparation for Future Learning (PFL) items asked students to apply a modified version of the taught procedure (Bransford & Schwartz, 2001). For example, after learning to calculate variability by averaging the errors (MD), students were asked to apply MD while ignoring certain types of errors (e.g., include only large errors, see Figure 2).
Debugging items asked students whether a modified version of the taught procedure still achieves its goal, and if it fails, to identify the source of failure (Roll, 2009).

The effect of condition on students' learning was evaluated using ANCOVAs with condition, class level, and their interaction as factors, and pre-test score as a covariate. Five separate ANCOVAs were calculated, one for each type of items. Bonferroni correction was applied to account for multiple comparisons.

Results

There was a large, yet insignificant, difference between conditions at pre-test: *Design and Evaluate*: 15%; *Evaluate Only*: 27%; $F(2,88)=10$, $p = .19$. On identical items in the pre- and post-tests, students improved from 23% to 58% across conditions and class levels; $t(90) = 7.0$, $p < .0005$.

Table 1 summarizes the results of the posttest. There was no effect for condition on procedural fluency items ($p = .21$). There was a significant main effect for condition on performance on conceptual items; *Design and Evaluate* = .52, *Evaluate Only* = .49, $F(4,87) = 18$, $p < .0005$. There was also a main effect for condition on debugging items; *Design and Evaluate*: .44, *Evaluate Only*: .34. $F(4,87) = 42$, $p = .002$. Both effects remain significant after applying Bonferroni correction. There was no effect for condition on PFL items with or without resource ($p =$ .2 and .3 respectively).

Table 1: Performance on post-test by type of item, M(SD)

		Design and Evaluate	*Evaluate Only*
Procedural Fluency		.64 (.44)	.62 (.45)
Conceptual understanding		.52 (.25)***	.49 (.24)
Debugging		.44 (.30)**	.34 (.34)
PFL	W/o resource	.09 (.20)	.06 (.16)
	With resource	.46 (.45)	.40 (.45)

** - $p < .01$; *** - $p < .001$

Discussion

Overall, students who designed their methods outperformed students who received predesigned methods on measures of debugging ability and conceptual understanding. There was no effect of condition on procedural fluency or PFL items.

Invention vs. Evaluation

The results presented above show the importance of students designing their own methods, and thus support Hypothesis 1(a). This effect is especially interesting given that all students failed to design valid methods.

There are several possible explanations for the benefits of design. First, students who design their methods have greater agency. Second, students may learn better from failures of methods they designed since they understand the intended function of each component in their methods. Third, students who work with pre-designed methods might have an overly specific goal – to test given methods using given data. Students who design methods may have a more scientific goal to understand the features of the domain and create a method that captures them. More studies are required to evaluate these explanations.

Knowledge Outcomes and Mechanisms of Transfer

Performance on conceptual items shows that students who design their methods acquire better domain knowledge, in support of Hypothesis 2(a). Their improved performance on debugging items might suggest that students who design also acquired better domain-independent debugging skills, as suggested by Hypothesis 2(b). However, students in both conditions practiced the relevant debugging skills during invention. In fact, Evaluate Only students had more time to practice debugging on tasks isomorphic to the assessment items. Therefore, a more likely interpretation is that the improved conceptual understanding, and not improved debugging skills, helped students who designed their methods perform better on debugging items.

Accumulated evidence from this and previous studies shows that invention activities do not help students *apply* the canonical methods (in procedural items), yet they help students *modify* the learned methods (in PFL items) and *diagnose variants* of these methods (in debugging items; Roll, 2009; Roll, et al., 2009; Schwartz & Martin, 2004). A main question to ask, therefore, is how the experience of failing to invent prepares students to modify and adapt their acquired knowledge following instruction.

As described above, students in invention activities invent using contrasting cases. The contrasting cases are designed to direct students' attention to deep features of the domain, often one feature at a time. In order for their methods to work, students need to find mathematical ways to represent the target features in their methods. For example, the contrasting cases shown in Figure 1 help students realize that their methods should use all available data, since relying only on extreme points may give the wrong answer. Thus, one explanation for the benefits of failed invention attempts is that *the invention process is essentially a process in which students acquire requirements for a valid method* (see Figure 3.a). Within the domain of variability, by using contrasting cases, students may realize that a valid method should measure distances, use all given data, account for sample size, and use positive values. Students may find mathematical operations that satisfy some, but not all, of these requirements. For example, few methods included division by N. In fact, most requirements are left unsatisfied.

The invention process helps students identify concrete, and possibly explicit, requirements, and examples of successful and unsuccessful ways to achieve them. Later, during show-and-practice, students can complete the puzzle by noticing how the canonical method satisfies these requirements. *Comprehension of instruction involves a mapping process in which students can identify how each component of the canonical method fills a certain function.* In the case of MD, division by N accounts for sample size; the minus operator

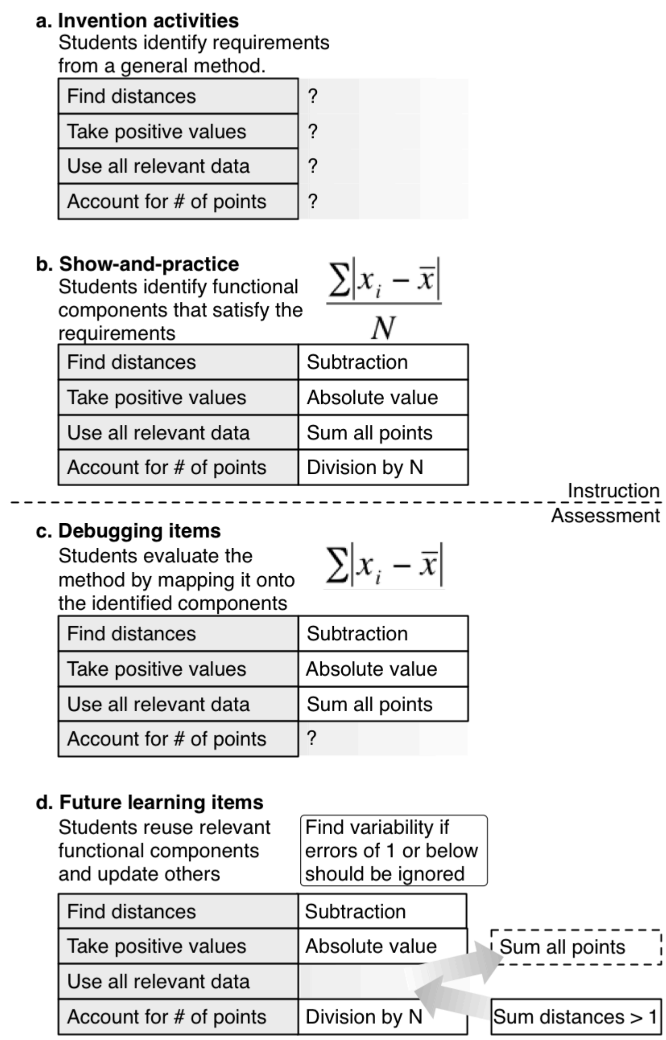

Figure 3: A possible explanation for learning and transfer from invention activities. While inventing, students can identify requirements from valid solutions (a). During show-and-practice students realize how the taught formalism addresses these requirements (b). Such functional encoding helps students adapt their knowledge (c, d)

calculates distances; the sum function uses all the points; and absolute value makes sure that distances do not cancel each other out. The outcome is knowledge that is more elaborated and is composed of critical functional components, each of which are backed up by the relevant rationale and supporting experiences.

Students in the Evaluate Only condition have some of these experiences, but these are not customized to their own preconceptions. More importantly, these students focus on evaluating methods, and not on identifying requirements.

The modular encoding of functional components (i.e., relating each component in the method to its specific function) allows students to adapt their knowledge to the task. For example, in debugging items, students can map the debugged procedures onto the set of requirements they have come to understand and identify which requirement is not met.

Summary

We present a study that compares invention activities that include design and evaluation of methods to evaluation of predesigned methods, prior to receiving instruction. Our results show that the design of methods enhances the effect of invention activities. The results also show that students who design their methods acquire better domain knowledge, and not necessarily better invention skills. Further analysis suggests that the invention process helps students set requirements from the general method. Students later identify functional components that satisfy these requirements within the canonical method, resulting in more differentiated and modular knowledge. The acquired knowledge can be applied flexibly by reusing and recombining the functional components.

The study makes contributions by identifying the cognitive benefits of invention activities and by explaining how these benefits interact with the taught material to improve the ability to transfer the target knowledge. Students who invent gain the knowledge of key requirements of formalisms, of reasons for these requirements, and of mathematical tools that satisfy (and fail to satisfy) these requirements.

Acknowledgements

This work is supported by the Pittsburgh Science of Learning Center, National Science Foundation award SBE-0354420.

References

Aleven, V., McLaren, B., Sewall, J., & Koedinger, K. R. (2009). Example-tracing *tutors*: A new paradigm for intelligent tutoring systems. *International Journal of Artificial Intelligence in Education, 19*: 105-154.

Bransford, J. D., & Schwartz, D. L. (2001). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education*, 24(3), 61-100.

Chase, C. C., Shemwell, J. T., & Schwartz, D. L. (2010). Explaining across contrasting cases for deep understanding in science: An example using interactive simulations. In *proceedings of the International Conference of the Learning Sciences 2010.*

Kapur, M. (2008). Productive failure. *Cognition and Instruction*, *26*(3), 379 - 424.

Richland, L. E., Bjork, R. A., Finley, J. R., & Linn, M. C. (2005). Linking cognitive science to education: Generation and interleaving effects. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th annual conference of the cognitive science society.* Mahwah, NJ: Lawrence Erlbaum.

Roll, I. (2009). *Structured invention activities to prepare students for future learning: Means, mechanisms, and cognitive processes.* Ph.D. Thesis, Pittsburgh, PA.

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*. Austin, TX: Cognitive Science Society.

Roll, I., Aleven, V., & Koedinger, K. R. (2010). The invention lab: Using a hybrid of model tracing and constraint-based modeling to offer intelligent support in inquiry environments. In V. Aleven, J. Kay, & J. Mostow (Eds.), *Proceedings of the international conference on intelligent tutoring systems.* Berlin: Springer Verlag.

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

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

Schwartz, D. L., Sears, D., & Chang, J. (2007). Reconsidering prior knowledge. In M. C. Lovett, & P. Shah (Eds.), *Thinking with data.* (pp. 319-44). New York: Routledge.

Sweller, J. (2006). The worked example effect and human cognition. *Learning and Instruction*, *16*(2), 165-169.

Taylor, J. L., Smith, K. M., van Stolk, A. P., & Spiegelman, G. B. (2010). Using invention to change how students tackle problems. *CBE-Life Sciences Education*, *9*(4).