## Title
How Much Support Is Optimal During Exploratory Learning?



## Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 40(0)

## Authors
Newman, Phillip M
DeCaro, Marci S

## Publication Date
2018

# How Much Support Is Optimal During Exploratory Learning?

**Phillip M. Newman (pmnewm03@louisville.edu)**
Department of Psychological and Brain Sciences, 2301 S. 3rd Street
Louisville, KY 40292 USA

**Marci S. DeCaro (marci.decaro@louisville.edu)**
Department of Psychological and Brain Sciences, 2301 S. 3rd Street
Louisville, KY 40292 USA

**Abstract**

Students who explore a new concept prior to receiving direct instruction often demonstrate better conceptual understanding compared to traditional tell-then-practice methods. Often, exploratory learning activities have students invent solutions to a novel problem targeting the new concept. However, exploring prior to instruction is working memory demanding, inducing high cognitive load. The current experiments varied the guidance provided during exploration and examined subsequent learning. In Experiment 1, participants explored the procedures and concept of statistical variance prior to receiving instruction in one of three conditions: invention, completion problem, or worked example. Exploring using a worked example led to the highest learning outcomes and the least cognitive load. In Experiment 2, students in an undergraduate statistics class completed invention or worked example problems either before or after instruction. Learning was greater when problem solving preceded instruction. However, exploring using a worked example did not improve learning over the more cognitively-demanding invention problem. These findings demonstrate the benefits of exploratory learning in the classroom compared to more traditional tell-then-practice approaches. However, more research is needed to determine when and how guidance will enhance exploration.

**Keywords:** exploratory learning; completion problems; worked examples; cognitive load; education

## Introduction

Typically, instructors directly teach mathematical problem-solving procedures and concepts, followed by problem-solving practice. An inverse approach, generally referred to as *exploratory learning* (DeCaro & Rittle-Johnson, 2012; Weaver, Chastain, DeCaro, & DeCaro, 2018), provides students an opportunity to explore a new concept prior to instruction. This approach has been shown to benefit conceptual understanding relative to traditional instruct-then-practice approaches (see Kapur, 2016; Loibl, Roll, & Rummel, 2016; Schwartz, Lindgren, & Lewis, 2009).

One specific method of exploratory learning is *learning-by-inventing* (LBI); students are asked to invent a method for solving a novel problem targeting the concept to be learned (Schwartz & Martin, 2004). Afterwards, students receive direct instruction. Importantly, exploratory learning methods such as LBI are not pure discovery learning, but combine aspects of both constructivist-inspired and direct instruction approaches. Previous studies have shown that LBI enhances students' understanding of concepts such as statistical variance and standard deviation (e.g., Schwartz & Martin, 2004; Wiedmann, Leach, Rummel, & Wiley, 2012).

## Cognitive Mechanisms Supporting LBI

LBI is thought to improve conceptual understanding through several key mechanisms. First, LBI helps students activate prior knowledge of relevant concepts (Kapur, 2012). This process enables students to prepare preexisting schemas in long-term memory to integrate new information from instruction (Sweller, Jeroen, & Paas, 1998).

Second, invention activities may improve metacognition, by helping students become aware of gaps between their current understanding and that required by the problem (Loibl & Rummel, 2014a). Awareness of knowledge gaps may increase interest and attention to subsequent instruction, by providing a "need to know" (Glogger-Frey et al., 2015; Loibl et al., 2016; Rotgans & Schmidt, 2014; Schwartz & Martin, 2004). In contrast, tell-then-practice methods lead students to perceive that they understand the material better than they actually do and decrease attention and effort (DeCaro & Rittle-Johnson, 2012; Renkl, 1999).

Third, LBI may enable students to recognize deep structural features of the problem (Loibl, Roll, & Rummel, 2016; Schwartz & Martin, 2004). Students must explore the problem space by testing hypotheses using a trial and error process (DeCaro & Rittle-Johnson, 2012). Students begin to determine which features are important for solving the problem, and which are not (Glogger-Frey et al., 2015; Loibl et al., 2016; Schwartz & Martin, 2004). This process supports deeper understanding.

## Rich Datasets

*Invention problems* typically help to achieve these key learning mechanisms by incorporating rich datasets (Loibl et al., 2016). One such problem used by Weidmann et al. (2012) asks students to invent a formula to calculate *consistency* for three small datasets. Following, students learn the statistical concepts and procedures of variance and standard deviation. Such problems encourage students to explore previously learned methods for analyzing data (e.g., calculating the mean, drawing bar/line graphs), activating prior knowledge. Additionally, because the mean of each dataset is equal, students cannot simply calculate the mean. When students reach an impasse like this, they become aware of gaps in their knowledge. Structural features are highlighted when students look further for similarities and differences between cases (Schwartz & Martin, 2004).

## The Case Against Invention

Although exploring a problem prior to instruction helps students develop basic experience with a new concept, minimally guided learning activities such as LBI have been criticized. Kirschner, Sweller, and Clark (2006) argue that requiring students to explore a large problem space taxes working memory resources critical for schema development. In addition, students rarely develop optimal solution approaches (Kapur, 2012). If students do not receive feedback, they may continue to use their suboptimal solutions on posttests following instruction (Kirschner et al., 2006; Sweller et al., 1998). In support of this critique, Glogger-Frey et al. (2015) and Likourezos & Kalyuga (2016) have shown that cognitive load is higher following LBI when compared with guided alternatives.

## An Alternative Approach to Invention

In response to the above criticisms, researchers have explored alternatives to LBI (e.g., Glogger-Frey et al., 2015; Likourezos & Kalyuga, 2016; Loibl & Rummel, 2014b). One such approach is to have students explore worked examples prior to instruction. *Worked examples* are problems for which completely worked-out solutions are provided, usually supplemented with brief explanations (cf. Glogger-Frey et al., 2015). However, worked examples in this context are not entirely akin to direct instruction. Studying worked examples prior to instruction allows learners to explore the conceptual bases of appropriate solution approaches vicariously by studying someone else's steps. Worked examples decrease working memory demand (i.e., cognitive load) by eliminating the problem space (Sweller et al., 1998). Previous research comparing worked examples with invention problems have reported mixed results, with some finding increased learning following worked examples (e.g., Glogger-Frey et al. 2015), and others showing comparable learning outcomes (e.g., Likourezos & Kalyuga, 2016).

# Experiment 1

By requiring learners to navigate a large problem space during exploration, invention problems may induce high cognitive load, potentially reducing learning. In Experiment 1, we examined the learning impact of providing more guidance during LBI. Using materials adapted from Weidmann et al. (2012), participants explored the concept of variance using a rich dataset, then received direct instruction. The level of guidance provided during exploration was manipulated in three conditions: pure invention (no guidance), completion problem (partial guidance), and worked example (full guidance). *Completion problems* looked exactly like worked examples, except that some items were left blank for participants to fill in. In this way, completion problems are a middle-ground between worked examples and unguided problems, because they reduce the problem space while enabling learners to generate their own solutions (Sweller et al., 1998). Completion problems have yet to be explored in a LBI context. In addition to measuring learning, we assessed perceptions of cognitive load, knowledge gaps, and interest.

We hypothesized that both completion problems and worked examples would reduce cognitive load and increase learning, compared to invention problems. We also explored whether completion problems would lead to better learning than worked examples. One possibility is that, by reducing cognitive load and eliciting generation of partial problem solutions, completion problems would lead to better learning than worked examples. Another possibility is that reducing cognitive load would be sufficient, and completion problems and worked examples would lead to comparable learning effects. We further hypothesized that perceived knowledge gaps and interest would be equal or higher in the invention condition (Glogger-Frey et al., 2015).

# Methods

## Participants

Undergraduate students ($N$=123; age $M$=19.02, $SD$=2.04; 63.6% female) participated for credit in an introductory psychology course. Participants were randomly assigned to one of three conditions: Invention ($n$=40), completion problem ($n$=42), or worked example ($n$=41). Four additional participants were excluded from analyses for failure to complete the posttest.

## Materials

**Pretest** Two items measured prior knowledge of statistics. A central tendency problem asked participants to find the mean, median, and mode of an array of numbers (adapted from Paas, 1992). A variance problem provided a table of cinema attendance data and asked participants to determine mathematically which of two cinemas enjoys the most consistent attendance (adapted from Kapur, 2012).

**Problem-Solving Activity** The problem-solving activity (adapted from Weidmann et al., 2012) asked participants to help a group of managers determine which of three tea growers produces tea with the most consistent levels of antioxidants. A table listed antioxidant levels for each tea grower over the past six years. Participants in the *invention condition* were instructed to invent a formula to calculate consistency for each tea grower to determine which grower produces the most consistent levels of antioxidants. Participants were instructed to complete the calculations and to decide on the most consistent tea grower. Participants in the *worked example condition* received the same problem with standard deviation completely worked out for each tea grower along with brief explanations for the calculations. The tea grower with the most consistent levels of antioxidants was circled, and participants were instructed to study the calculations. The completion problem condition was the same as the worked example condition, with some blanks for participants to fill in.

**Questionnaire** Cognitive load was measured with the Mental Effort Rating Scale (Paas, 1992; "In solving or studying the previous problem I invested…"). Participants responded on a scale from 1 (very, very low mental effort) to 9 (very, very high mental effort). Interest was measured with three items (McDonald's ω=.88) adapted from Ryan's (1982) Intrinsic Motivation Inventory (e.g., "I found this learning activity interesting.") Perceived knowledge gaps were measured with four items (McDonald's ω=.89) adapted from Flynn and Goldsmith (1999; e.g., "I do not feel very knowledgeable about calculating consistency.") Interest and perceived knowledge gaps were rated on a 5-point Likert scale (1= strongly disagree; 5=strongly agree).

**Instruction** The direct instruction was provided in a text passage adapted from Weidmann et al. (2012). Participants were told that engineers were interested in comparing which trampoline (A or B) has the most consistent levels of bounciness. A table displaying data for inches of rebound for trampoline A was displayed followed by the canonical formula and step-by-step instructions for how to calculate standard deviation. Text-boxes explained concepts and defined mathematical calculations. A table displaying inches of rebound for Trampoline B was then presented, followed by three questions to help participants practice and further develop their understanding of standard deviation.

**Posttest** The posttest measured procedural fluency (1 item), conceptual understanding (2 items), and transfer (1 item). Items were drawn from Weidmann et al. (2012) and a psychological statistics exam. All items were scored on a four-point scale. Twenty percent of the items were scored by a second observer (interrater reliability: $r$=.90).

## Procedure

Participants were run in sessions of up to fifteen in a reserved classroom. After providing consent, participants were instructed that they would be learning about variance in statistics. Participants were provided with a standard calculator and completed an individual differences questionnaire and pretest (8 min). The questionnaire was administered as part of a larger study and will not be discussed further.

Afterwards, participants worked individually the problem-solving activity (15 min). Packets were interleaved by condition, and participants were randomly assigned to condition based on which packet they received. Following, participants completed the questionnaire and instruction (15 min). Then participants completed the posttest (30 min). Finally, participants were debriefed.

## Results

**Preliminary Analyses** Pretest items were examined as a function of condition, revealing no effect on the central tendency item (invention: $M$=2.04, $SD$=0.92; completion problem: $M$=2.31, $SD$=0.87; worked example: $M$=2.50, $SD$=0.74), $F(2,120)=1.52$, $p$=.224. However, condition had a significant effect on the variance item, (invention: $M$=1.10, $SD$=0.65; completion problem: $M$=1.52, $SD$=0.67; worked example: $M$=1.02, $SD$=0.80), $F(2,120)=5.75$, $p$=.004. Despite random assignment, prior knowledge of variance was unequal across conditions. Thus, this variable was used as a covariate in all subsequent analyses.

**Learning Outcomes** Posttest scores were examined using a 3 (condition: invention, completion problem, worked example) × 3 (posttest subscale: procedural, conceptual, transfer) repeated measures ANCOVA, with condition as a between-subjects factor and posttest subscale as a within-subjects factor. The assumption of sphericity was violated, $p$=.017. Therefore, the lower-bound statistic was used. A main effect of posttest subscale was found, $F(1,119)=16.05$, $p<.001$, $\eta_p^2=.14$. Post-hoc comparisons with Bonferroni correction (α=.016) revealed that students scored higher on procedural ($M$=3.30, $SD$=0.94) compared with conceptual ($M$=2.23, $SD$=1.10) and transfer ($M$=2.41, $SD$=1.03) subscales, $p<.001$. Conceptual and transfer subscales did not differ significantly, $p$=.094. A marginally-significant main effect of condition was found, $F(2,119)=3.01$, $p$=.053, $\eta_p^2=.05$ (see Figure 1). Planned comparisons revealed that the completion problem ($M$=2.71, $SD$=0.85) did not improve posttest performance compared to invention ($M$=2.40, $SD$=0.85), $p$=.098. However, the worked example ($M$=2.83, $SD$=0.73) led to significantly higher posttest performance than invention, $p$=.019. There was no interaction between condition and posttest subscale, $F$<1, indicating that the effects of condition occurred across the subscales.

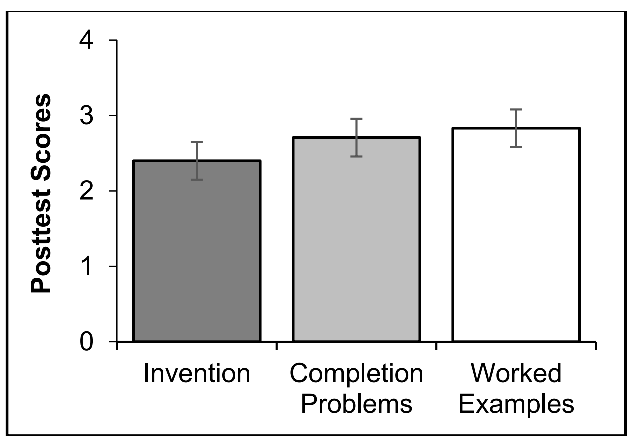


*Figure 1*: Posttest scores as a function of condition. Error bars represent 95% confidence intervals.

**Questionnaires** A significant effect of condition was found for cognitive load, $F(2,118)=3.20$, $p$=.045, $\eta_p^2=.05$. Planned comparisons demonstrated that completion problems ($M$=4.98, $SD$=2.02) did not lead to significantly less cognitive load than invention ($M$=5.70, $SD$=0.99), $p$=.063. Worked examples ($M$=4.80, $SD$=1.86) led to significantly less cognitive load than invention, $p$=.019.

A significant effect of condition was found for perceived knowledge gaps, $F(2,117)=13.83$, $p<.001$, $\eta_p^2=.19$. Planned comparisons revealed that completion problems ($M=2.79$, $SD=1.09$) and worked examples ($M=2.85$, $SD=0.93$) led to significantly lower perceived knowledge gaps than invention ($M=3.90$, $SD=0.80$), $p$s<.001.

Interest did not differ as a function of condition (Invention: $M=3.13$, $SD=0.91$; Completion problem: $M=3.43$, $SD=0.84$; Worked example: $M=3.20$, $SD=.95$), $F(2,116)=1.62$, $p=.203$, $\eta_p^2=.02$.

## Discussion

Participants who studied a worked example prior to instruction outperformed their inventing counterparts on the posttest, replicating the effects found by Glogger-Frey et al. (2015). Additionally, studying a worked example led to significantly less cognitive load than inventing. In contrast, participants in the completion problem condition did not outperform those in the invention condition, and also rated similar cognitive load. These results suggest that, despite allowing for generation, completion problems may be a less viable alternative to inventing compared to studying a worked example prior to instruction.

We did not find support for the notion that knowledge gaps experienced during minimally guided activities enhance learning from subsequent instruction. Although participants in the invention condition experienced the greatest knowledge gaps, they showed the poorest learning outcomes. Furthermore, we did not find support for the idea that the invention problem increases interest relative to the other exploratory learning conditions. Despite being more proscribed, worked examples did not reduce interest.

## Experiment 2

Previous studies have only explored guided alternatives to inventing within an exploratory learning context (Glogger-Frey et al., 2015; Likourezos & Kalyuga, 2016). Experiment 2 utilized a 2 (activity: invention, worked example) × 2 (order of instruction: explore-first, instruct-first) factorial design to compare the use of worked examples and invention problems both before and after instruction. Experiment 2 also attempted to replicate the findings from Experiment 1 in a psychological statistics course.

We hypothesized that those in the explore-first conditions would outperform those in the instruct-first conditions on the posttest. We also hypothesized that those who explored using a worked example would outperform those who invented prior to instruction. We also hypothesized that cognitive load would be highest for those who invented prior to instruction, compared to the other conditions.

## Methods

### Participants

Participants were 190 undergraduate students (Age $M=20.67$, $SD=4.33$; 72.9% female) enrolled in three sections of a psychological statistics course, across two semesters, with two different instructors of record. Participants were randomly assigned to one of four conditions: Explore-first/worked example ($n=46$), explore-first/invention ($n=48$), instruct-first/worked example ($n=47$), or instruct-first/invention ($n=49$). Additional participants were excluded from analyses for failure to provide consent ($n=3$), failure to complete the posttest ($n=11$), absence on the day of the posttest or inability to link their posttest to their first session packet (e.g., no name on the packet; $n=24$), or for having participated in Experiment 1 ($n=5$).

## Materials

The materials used in Experiment 2 were identical to those in Experiment 1, aside from three changes: (1) Because the variance problem on the pretest may serve as an invention activity itself (Kapur, 2016), the pretest was cut from the procedure; (2) A prompt in the worked example asked participants if they agreed with the chosen tea grower; (3) The consent form, problem-solving activity, questionnaire, and instruction were combined into one packet, with signals to stop and wait for instruction at the end of each section.

There were four different packets—one per condition. Invention and worked example conditions were the same as in Experiment 1. These problem-solving activities were provided either before instruction (explore-first conditions) or afterwards (instruct-first conditions). As in Experiment 1, 20% of posttests were scored by a second rater (interrater reliability: $r=.90$).

## Procedure

Participants completed the study across two lab sessions of their psychological statistics course. Both sessions occurred at the beginning of the semester, prior to lectures covering standard deviation and variance, and were 1-2 weeks apart. The first session included the problem-solving activity, questionnaire, and direct instruction. The second session included the posttest. Participants were provided with standard calculators in both sessions.

In the first session, participants were randomly assigned to condition based on the packet they received, which were interleaved. Following consent, participants completed the first section of the packet (problem-solving activity/ questionnaire or direct instruction, depending on condition; 15 min). Participants then completed the second section (problem-solving activity/questionnaire or direct instruction, depending on condition; 15 min). In the second session, participants completed the posttest (30 min) and were debriefed.

## Results

As described above, data were gathered from psychological statistics courses led by two different instructors. Because of possible differences between instructors, this variable was included as a covariate in all analyses.

**Learning Outcomes** Posttest performance was examined with a 3 (posttest subscale: procedural, conceptual, transfer) × 2 (order of instruction: explore-first, instruct-first) × 2 (activity: invention, worked example) ANCOVA, with posttest subscale as a within-subjects factor and order of instruction and activity as between-subjects factors. The assumption of sphericity was violated, $p<.001$, so the lower-bound statistic was used. A significant effect of posttest subscale was found $F(1,185)=32.74$, $p<.001$, $\eta_p^2=.14$. Post-hoc comparisons with Bonferroni correction ($\alpha=.016$) revealed that participants scored higher on procedural ($M=3.03$; $SD=1.12$) than conceptual ($M=2.21$; $SD=1.21$) and transfer subscales ($M=2.53$; $SD=1.19$), $p$s<.001. Transfer scores were higher than conceptual understanding, $p<.001$. Supporting our hypothesis, a main effect of order of instruction was found, with those in the explore-first condition outperforming their instruct-first counterparts, $F(1,185)=4.29$, $p=.040$, $\eta_p^2=.02$ (Figure 2). There was no main effect of activity or interaction, $F$s<1. A planned comparison revealed similar posttest scores for those who explored a worked example ($M=2.76$; $SD=1.10$) or invented ($M=2.73$; $SD=0.99$) prior to instruction, in contrast to our hypothesis, $p>.05$.

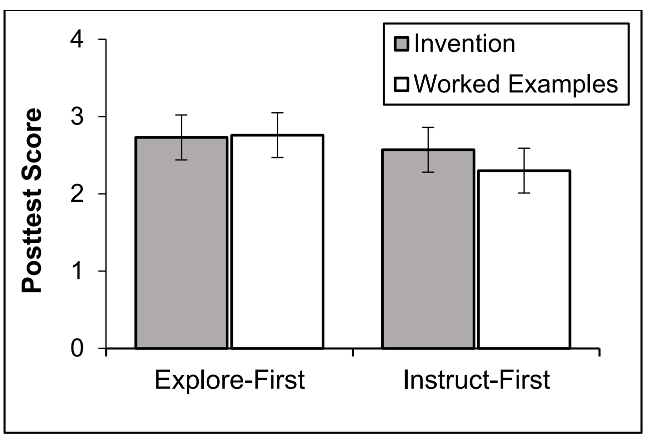


*Figure 2*: Posttest scores as a function of condition. Error bars represent 95% confidence intervals.

**Questionnaires** An ANCOVA revealed a significant main effect of activity on cognitive load, with those in the worked example conditions reporting less cognitive load ($M=4.91$; $SD=1.64$) than those in the invention conditions ($M=5.57$; $SD=1.49$), $F(1,175)=7.75$, $p=.006$, $\eta_p^2=.04$. Order of instruction did not significantly affect cognitive load, $F(1,175)=1.07$, $p=.302$, $\eta_p^2=.01$. There was no interaction, $F(1,175)=2.82$, $p=.095$, $\eta_p^2=.02$. Supporting our hypothesis, planned comparisons showed that, in the explore-first conditions, those who studied a worked example ($M=4.59$, $SD=1.61$) reported less cognitive load than those who invented ($M=5.65$, $SD=1.31$), $p=.001$. In the instruct-first conditions, cognitive load ratings after a worked example ($M=5.24$; $SD=1.64$) were comparable to inventing ($M=5.49$; $SD=1.69$), $p=.449$.

For perceived knowledge gaps, an ANCOVA revealed a significant main effect of order of instruction, with those in the explore-first conditions ($M=3.36$; $SD=1.05$) reporting greater knowledge gaps than those in the instruct-first conditions ($M=2.76$; $SD=0.96$), $F(1,175)=16.99$, $p<.001$, $\eta_p^2=.09$. There was also a main effect of activity, with those who invented ($M=3.33$; $SD=1.06$) reporting greater knowledge gaps than those who studied a worked example ($M=2.83$; $SD=1.05$), $F(1,175)=9.07$, $p=.003$, $\eta_p^2=.05$. These effects were qualified by a significant interaction, $F(1,175)=20.46$, $p<.001$, $\eta_p^2=.11$. Planned comparisons showed that those who invented prior to instruction ($M=3.88$; $SD=0.84$) perceived greater knowledge gaps than their worked example counterparts ($M=2.82$; $SD=0.97$), $p<.001$. Perceived knowledge gaps were similar for both activity groups that received instruction first, $p=.351$.

For interest, an ANCOVA revealed a non-significant main effect of order of instruction on interest, with those in the instruct-first condition ($M=3.50$; $SD=.88$) tending to report higher interest than those in the explore-first condition ($M=3.27$; $SD=.75$), $F(1,175)=3.38$, $p=.068$, $\eta_p^2=.02$. No main effect of activity, $F<1$, or interaction, $F(1,175)=1.25$, $p=.265$, $\eta_p^2=.01$, were found.

## General Discussion

In Experiment 1, learning outcomes were greatest for those who studied a worked example prior to receiving instruction. However, Experiment 2 did not replicate this finding, as both explore-first conditions showed comparable learning despite lower cognitive load in the worked example condition. There were two key differences between Experiments 1 and 2 that might account for these inconsistent findings. First, Experiment 1 was conducted with laboratory participants from an introductory psychology course, whereas Experiment 2 was conducted in the classroom, with more advanced students. Perhaps greater prior knowledge or motivation improved learning in the invention condition in the classroom sample.

Related to this point, Experiment 2 did not include a pretest, whereas Experiment 1 did. Thus, prior knowledge could not be accounted for in Experiment 2. On the other hand, the pretest in Experiment 1 may have actually bolstered the effect of studying a worked example, by serving as an invention activity (Kapur, 2016). The pretest asked participants to find the mean, median, and mode of a dataset, thus activating relevant prior knowledge. Additionally, the pretest included a variance problem asking students to determine mathematically which cinema enjoys the most consistent attendance, an item similar to the invention problem. Thus, participants in Experiment 1 may have engaged in important exploratory learning processes during the pretest (e.g., activating relevant prior knowledge, and attending to knowledge gaps and key problem features). It is possible that worked examples are optimal when preceded by a pretest because the learner is exposed to unguided exploration prior to guided exploration, receiving unique benefits from both: Recognition of knowledge gaps during invention, followed by a less cognitively demanding (guided) exploratory activity using a rich dataset in which optimal solutions are explored vicariously. Future research should explore the impact of using a pretest combined with varying levels of guidance during exploration.

Across both experiments, invention led to greater knowledge gaps but did not enhance interest relative to worked examples. This finding is inconsistent with literature demonstrating a link between knowledge gaps and situational interest (e.g., Rotgans and Schmidt, 2014). One possibility is that interest items asked about interest in the problem-solving activity, whereas knowledge gaps may enhance interest in the instruction and the subject in general, which was not captured by our measures. Additionally, given the classroom context, students may have shown a greater interest in activities for which they felt the most confident and familiar.

In conclusion, the current work replicates and extends previous research demonstrating the benefit of exploratory learning over traditional tell-then-practice methods in an undergraduate classroom context (Exp. 2). In addition, this research demonstrates that pure invention may not be necessary when designing exploratory learning materials. Using worked examples during exploration decreased cognitive load and resulted in equal (Exp. 2) or better (Exp. 1) learning outcomes than invention. This work suggests that exploration does not have to be difficult to be beneficial (cf. Kapur, 2016). By further examining the cognitive mechanisms by which various exploratory learning materials impact learning, researchers and educators can better understand when and why this method supports conceptual understanding.

## References


Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe, & A. Shimamura (Eds.). *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.

DeCaro, M. S., & Rittle-Johnson, B. (2012). Exploring mathematics problems prepares children to learn from instruction. *Journal of Experimental Child Psychology, 113*, 552-568.

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction, 22(4),* 271–280. http://dx.doi.org/10.1016/j.learninstruc.2011.08.003.

Flynn, L. R., & Goldsmith, R. E. (1999). A short, reliable measure of subjective knowledge. *Journal of Business Research, 46*(1), 57-66. doi:10.1016/S0148-2963(98)00057-5.

Glogger-Frey, I., Fleischer, C., Grüny, L., Kappich, J., & Renkl, A. (2015). Inventing a solution and studying a worked solution prepare differently for learning from direct instruction. *Learning and Instruction, 39*, 72-87.

Kapur, M. (2012). Productive failure in learning the concept of variance. *Instructional Science, 40*(4), 651-672.

Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. Educational Psychologist, 51(2), 289–299. http://dx.doi.org/10.1080/00461520.2016.1155457.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist, 41*(2), 75-86. doi:10.1207/s15326985ep4102_1.

Likourezos, V., & Kalyuga, S. (2016). Instruction-first and problem-solving-first approaches: Alternative pathways to learning complex tasks. *Instructional Science*, doi:10.1007/s11251-016-9399-4.

Loibl, K., Roll, I., & Rummel, N. (2016). Towards a theory of when and how problem solving followed by instruction supports learning. *Educational Psychology Review*, doi:10.1007/s10648-016-9379-x.

Loibl, K. & Rummel, N. (2014a). Knowing what you don't know makes failure productive. *Learning and Instruction, 34*, 74–85.

Loibl, K., & Rummel, N. (2014b). The impact of guidance during problem-solving prior to instruction on students' inventions and learning outcomes. *Instructional Science, 42(3)*, 305-326.

Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84*(4), 429-434. doi:10.1037/0022-0663.84.4.429.

Renkl, A. (1999). Learning mathematics from worked-out examples: Analyzing and fostering self-explanations. *European Journal of Psychology of Education, 14(4)*, 477–488.

Rotgans, J. I., & Schmidt, H. G. (2014). Situational interest and learning: thirst for knowledge. *Learning and Instruction, 32*, 37–50. http://dx.doi.org/10.1016/j.learninstruc.2014.01.002.

Ryan, R. M. (1982). Control and information in the intrapersonal sphere:An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, *43*, 450-461.

Schwartz, D. L., Lindgren, R., & Lewis, S. (2009). Constructivism in an age of non-constructivist assessments. In S. Tobias, & T. M. Duffy (Eds.). *Constructivist instruction: Success or failure?* (pp. 34–61). New York, NY: Routledge/Taylor & Francis Group.

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.

Sweller, J., van Merrienboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296.

Wiedmann, M., Leach, R. C., Rummel, N., & Wiley, J. (2012). Does group composition affect learning by invention? *Instructional Science, 40*(4), 711–730.