Learning and Instruction 62 (2019) 49–63

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

Learning by exploring: How much guidance is optimal?

Phillip M. Newman[1], Marci S. DeCaro*

Department of Psychological and Brain Sciences, 317 Life Sciences Building, University of Louisville, Louisville, KY, 40292, USA

ARTICLE INFO

Keywords:
Exploratory learning
Inventing
Productive failure
Worked examples
Cognitive load
Knowledge gaps

ABSTRACT

Exploring a new concept before instruction can benefit conceptual understanding, but is demanding. The current experiments examined whether providing guidance during exploration improves learning. Undergraduate students explored the procedures and concept of statistical variance prior to direct instruction. In Experiment 1 (N = 123), exploring using worked examples (full guidance) led to higher posttest scores than exploring using an invention activity (no guidance) or completion problems (partial guidance). In contrast, Experiment 2 (N = 190) found no learning benefit of exploring using worked examples compared to inventing. Overall, exploring improved learning compared to instruct-then-practice conditions. Experiment 3 (N = 147) demonstrated that exploring worked examples improved learning compared to exploring using an invention activity—but only when preceded by a pretest. Students' reported cognitive load, knowledge gaps, and interest were also assessed. Findings suggest that combining a pretest with worked examples helps students perceive knowledge gaps and discern problem features, maximizing exploration while reducing cognitive load.

1. Introduction

Traditionally, instructors first teach concepts and procedures, followed by problem-solving practice. However, this approach may lead to superficial attention and learning (DeCaro & Rittle-Johnson, 2012; Renkl, 1999). Students do not always understand how the new information fits in with what they already know, and do not meaningfully integrate the new information with prior knowledge (Schwartz & Bransford, 1998). Students also focus on superficial features rather than deeper principles (Glogger-Frey, Gaus, & Renkl, 2017). Additionally, the ease of listening to a lecture can lead students to believe they understand the material better than they do (Bjork, 1994; Renkl, 1999). For these reasons, the traditional approach often results in weak conceptual understanding and limited ability to transfer to new contexts (Kapur, 2012; Loibl & Rummel, 2014a, 2014b).

To counter these concerns, some instructors select constructivist-inspired methods that offer minimal guidance, such as discovery learning (Alfieri, Brooks, Aldrich, & Tenenbaum, 2011; Dean & Kuhn, 2007). Students are responsible for discovering underlying patterns, absent of instructional guidance. However, purely constructivist approaches ignore capacity limits of working memory (Kirschner, Sweller, & Clark, 2006), and often lead to weaker learning relative to guided instruction (cf. Alfieri et al., 2011; Mayer, 2004).

Other approaches draw on both traditional and constructivist-inspired instructional methods, often resulting in better learning than either approach alone (Alfieri et al., 2011). For example, *exploratory learning* includes two phases. First, students explore a new concept, in keeping with constructivist-inspired methods. Then, students are given instruction (DeCaro & Rittle-Johnson, 2012; Schwartz, Lindgren, & Lewis, 2009; Weaver, Chastain, DeCaro, & DeCaro, 2018). We use the term *exploratory learning* to encompass several specific literatures examining this two-phase procedure (i.e., inventing to prepare for future learning, productive failure, and problem-solving first methods; Likourezos & Kalyuga, 2016; Loibl, Roll, & Rummel, 2016). Typically, procedures are learned equally well with traditional and exploratory learning methods. However, conceptual understanding and transfer have been shown to improve when using exploratory learning approaches compared to traditional instruction (see Kapur, 2015, 2016; Schwartz et al., 2009).

The types of exploration activities differ across studies (Loibl et al., 2016). For example, one type of exploratory learning activity, *invention problems*, ask students to invent a method for solving a novel problem targeting the to-be-learned concept (Schwartz & Martin, 2004). Afterwards, students receive direct instruction. Previous studies have shown that invention with subsequent instruction enhances students' understanding of concepts such as statistical variance and standard deviation (d = 1.29–2.51, η_p^2 = 0.08 − 0.30; e.g., Jarosz, Goldenberg, & Wiley, 2016; Kapur, 2012, 2014; Loibl & Rummel, 2014a, 2014b; Schwartz &

* Corresponding author.
E-mail address: marci.decaro@louisville.edu (M.S. DeCaro).
[1] Phillip M. Newman is now in the Department of Psychology, Vanderbilt University, Nashville, TN.

https://doi.org/10.1016/j.learninstruc.2019.05.005
Received 1 November 2018; Received in revised form 15 May 2019; Accepted 17 May 2019

Martin, 2004). For example, Kapur (2012) found that ninth-grade students who invented methods for calculating variance prior to instruction demonstrated greater conceptual understanding and transfer on a posttest than students who received instruction followed by practice.

Across the current studies, we manipulate the amount of guidance students receive during an invention activity prior to direct instruction. We also examine whether including a pretest changes the benefits of exploring. We assess students' self-reported cognitive load, knowledge gaps, and interest/enjoyment. By comparing how different types of invention activities impact learning and self-reports, we can better understand when and why exploratory learning will support students' understanding.

1.1. Cognitive mechanisms supporting exploratory learning

Several mechanisms have been proposed to explain how exploration activities improve learning (Kalyuga & Singh, 2016; Loibl et al., 2016). First, exploratory learning likely has metacognitive and motivational benefits, by helping students become aware of gaps between their current understanding and that required by the problem (Glogger-Frey, Fleischer, Grüny, Kappich, & Renkl, 2015; Loibl & Rummel, 2014a). *Awareness of knowledge gaps* may increase interest and help guide attention during subsequent instruction by providing an impasse that must be resolved, or a "need to know" (Berlyne, 1954; Glogger-Frey et al., 2015; Glogger-Frey et al., 2017; Richland, Kornell, & Kao, 2009; Rotgans & Schmidt, 2014; Schwartz & Martin, 2004; Wise & O'Neill, 2009). Situational *interest and enjoyment* can also be engaged when a task is novel, somewhat complex, and requires personal direction (Rotgans & Schmidt, 2014; Silvia, 2008). In contrast, traditional instruct-then-practice methods are less likely to lead students to perceive knowledge gaps, creating an illusory perception that they understand the material better than they actually do, and decreasing attention and effort (Bjork, 1994; DeCaro & Rittle-Johnson, 2012; Dunlosky & Rawson, 2012; Kapur, 2016; Renkl, 1999).

Second, exploratory learning is posited to have cognitive benefits. Engaging in an exploration activity enables students to *activate prior knowledge* of relevant concepts, such as central tendency when learning about variance (Kapur, 2012, 2015; Schwartz, Sears, & Chang, 2007). Prior knowledge is stored in long-term memory as schemas (Sweller, 2004; van Merriënboer & Sweller, 2005). Activating prior knowledge during exploratory learning allows students to prepare preexisting schemas to integrate with new information from instruction (Sweller, van Merriënboer, & Paas, 1998). In addition, exploratory learning is thought to help students better *perceive deep structural problem features* (Kapur, 2012; Loibl et al., 2016; Schwartz & Bransford, 1998; Schwartz & Martin, 2004). Students must test solution possibilities using trial and error (DeCaro & Rittle-Johnson, 2012). Students then begin to determine which features are important for solving the problem, and which are not (Glogger-Frey et al., 2015; Schwartz & Martin, 2004). This process is thought to support deeper understanding and transfer to problems containing similar structural features (Glogger-Frey et al., 2017; Schwartz & Bransford, 1998).

1.2. The case against exploratory learning

Despite evidence that exploratory learning is beneficial, these activities can be difficult. Students must select among many possible solution paths, leading to errors or failure to reach a solution (Kapur, 2016; Kirschner et al., 2006). Drawing on cognitive load theory, Kirshner, Sweller, and Clark (2006) argue that such activities harm learning, because they induce high *cognitive load.* Students' limited working memory resources are directed towards inappropriate solution approaches rather than correct information (see also Hsu, Kalyuga, & Sweller, 2015; Mayer, 2004; Mayer & Moreno, 2003). As described by Kalyuga and Singh (2016), cognitive load theory states that learning activities should maximize intrinsic load while reducing extraneous load. Both intrinsic and extraneous load result in mental effort, but intrinsic load is productive and extraneous load is unproductive for attaining domain-specific knowledge (Kalyuga & Singh, 2016). Consistent with this criticism, research has demonstrated that cognitive load is higher following invention activities when compared with guided alternatives (Glogger-Frey et al., 2015, $d = 0.94$; Likourezos & Kalyuga, 2016, $\eta_p^2 = 0.15$).

Kalyuga and Singh (2016) recently outlined how exploratory learning findings may be reconciled with cognitive load theory, concluding that cognitive load theory does not apply to exploration activities. Specifically, Kalyuga and Singh outline three general goals of learning: (1) to prepare students to learn, for example by activating knowledge gaps or situational interest prior to instruction; (2) to develop domain-specific knowledge, such as learning specific facts and procedures; and (3) to develop general concepts and the ability to transfer within a domain. Kalyuga and Singh state that cognitive load theory has traditionally focused only on the second goal, but may be relevant to the third goal as well. However, they posit that cognitive load is less relevant for the first goal, which includes exploratory learning, because exploration is not intended to instruct students on specific facts and procedures. They acknowledge that allowing learners to encounter multiple solution paths and make errors may benefit later learning, despite inducing cognitive load. In addition, they note that cognitive load may still impact whether exploration activities successfully prepare students for future learning.

1.3. Providing guidance during exploratory learning

In the current research, we directly test whether cognitive load matters during exploratory learning. Kalyuga and Singh (2016) emphasize the metacognitive and motivational processes of exploratory learning, but largely overlook the cognitive benefits. If cognitive load during exploration is too high, learners may be less likely to perceive relevant problem features. Thus, high cognitive load—even during an exploration activity—may impact conceptual understanding and transfer.

One way to test this idea is to examine how providing different levels of guidance during exploration impacts learning. Initial studies have compared exploration using an invention activity to exploration using worked examples prior to instruction. *Worked examples* are problems for which completely worked-out solutions are provided, usually with brief explanations (cf. Chen, Kalyuga, & Sweller, 2015; Glogger-Frey et al., 2015, 2017). Worked examples given in an exploratory learning context allow learners to explore the conceptual bases of appropriate solution approaches by studying someone else's steps. Worked examples decrease cognitive load by eliminating multiple solution paths and errors (Sweller et al., 1998; Tuovinen & Sweller, 1999). Worked examples provided prior to direct instruction are considered to be a form of guided exploration, as students are guided in the procedures but not necessarily the underlying concepts. Students are preparing to learn the full information provided in the subsequent instruction (Glogger-Frey et al., 2015, 2017).

Previous research comparing worked examples to invention activities prior to instruction have reported mixed results. Across two experiments, Glogger-Frey et al. (2015) showed enhanced learning for students who explored using worked examples compared to those given an invention activity ($d = 0.71 - 0.72$)—even though inventing led to higher perceived knowledge gaps, epistemic curiosity, and interest. In the first experiment, student teachers learned how to evaluate the quality of student learning journals. In the second experiment, eighth-grade students learned about density and ratio indices in physics. Cognitive load was also assessed and was lower in the worked-examples condition ($d = 0.94$).

Likourezos and Kalyuga (2016) compared three levels of guidance during an exploration task: No guidance, partial guidance (the final solution was provided, guiding participants to the correct solution), and

worked examples. Year 8 high school students learned about properties of geometric figures. No differences were found between conditions on the posttest measures. However, students in the worked examples condition generated more creative and advanced responses during the posttest. Students who explored worked examples also reported lower extraneous load ($\eta_p^2 = 0.15$). These results provide limited support for the idea that reducing cognitive load helped students in the worked examples condition develop more sophisticated schemas.

In contrast to these studies, Glogger-Frey et al. (2017) found reduced cognitive load, but also lower learning, in a worked examples condition, compared to an invention condition in which students received extra practice (far-transfer: $d = 0.61$; near-transfer: $d = 0.54$). Participants were eight-grade students learning about density and ratio indices in physics.

2. Current studies

The current studies further examined whether providing guidance during exploration benefits learning. Across three experiments, undergraduate students were taught a different topic than in the previous studies, the concept of variance, using an activity adapted from Wiedmann, Leach, Rummel, and Wiley (2012).

In Experiment 1, students in a laboratory setting completed one of three exploratory learning conditions that varied in the support provided. Students were provided with a table containing three datasets, representing the amount of antioxidants in tea produced by three tea growers over several years. Students in the *invention condition* were asked to invent a method to evaluate which tea grower produced the most consistent level of antioxidants. Students in the *worked examples condition* had the method of calculating consistency worked out for them, with explanations for each step (see Fig. 1). Students in the *completion problems condition* viewed the same worked examples, but with some blanks for them to fill in (Fig. 1). Following the learning activity and instruction, all students completed a posttest measuring procedural fluency (ability to calculate standard deviation), conceptual understanding, and transfer.

Completion problems were intended to limit the solution possibilities while also enabling learners to generate some solutions (Chen et al., 2015; Paas, 1992; Slamecka & Graf, 1978; Sweller et al., 1998). Although previous research has examined partially guided learning activities prior to instruction, these activities either provided the final solution without revealing the steps (e.g., Likourezos & Kalyuga, 2016), or provided guidance with contrasting cases (e.g., Loibl & Rummel, 2014b). Completion problems in which learners complete partially worked out solutions have yet to be explored in an exploratory learning context.

Experiments 2 and 3 further compared the use of worked examples to invention problems during exploratory learning. Experiment 2 compared these conditions to instruct-then-practice conditions in undergraduate statistics courses. This experiment allowed us to examine whether findings generalize to the classroom and whether using worked examples as exploration had greater learning benefits than receiving instruction first. Experiment 3 tested a potential boundary condition—whether the use of a pretest enhances the benefits of using worked examples during exploration. Pretests may serve as a form of exploration activity (e.g., activating prior knowledge and perceptions of knowledge gaps; Glogger-Frey et al., 2015, 2017; Kapur, 2016). However, this possibility has never been systematically tested in the exploratory learning literature. Use of a pretest has also not been tested in conjunction with worked examples. Together, these instructional elements may provide a synergistic mixture of both guidance and exploration.

Thus, the current studies further test the potential benefits of designing exploration activities that take cognitive load into account. We extend prior studies by examining this question in a new domain across both laboratory and classroom settings, and by comparing the use of worked examples as exploration to instruct-then-practice conditions. We further build theory by examining the impact of providing a pretest prior to exploring. In all three studies, we assess learning outcomes and potential learning mechanisms, including students' perceived knowledge gaps, cognitive load, and situational interest and enjoyment. By doing so, these studies provide new insight into not only what factors improve learning from exploration, but why.

3. Experiment 1

In Experiment 1, we manipulated the level of guidance provided during exploratory learning in three conditions: invention (no guidance), completion problems (partial guidance), and worked examples (full guidance). Learning was assessed on an immediate posttest assessing procedural fluency, conceptual knowledge, and transfer. We measured students' perceived knowledge gaps, cognitive load, and interest and enjoyment following the activity. We did not differentiate between intrinsic and extraneous cognitive load in our measure. Kalyuga and Singh (2016) argued that these definitions of cognitive load do not apply to activities given prior to direct instruction. They instead define cognitive load for these activities as “the intensity of cognitive activity involved in achieving a specific goal of the task” (Kalyuga & Singh, 2016, p. 848). Consistent with this framing, we measured cognitive load as perceived mental effort during the activity, using a measure commonly implemented in education research (Paas, 1992; see also; Hsu et al., 2015). Of course, self-report measures rely on individuals' perceptions, and may not always reflect objective experience.

Although Kalyuga and Singh (2016) argued that cognitive load is not relevant to exploration activities, we hypothesized that reducing cognitive load during exploration would increase learning. Specifically, we tested the following predictions:

1. Learning Outcomes: We hypothesized that worked examples and completion problems would increase posttest scores, compared to invention. Comparing completion problems and worked examples, we expected one of two possible outcomes. One possibility is that, by reducing cognitive load and asking students to generate partial problem solutions, completion problems will enhance learning relative to worked examples. Another possibility is that reducing cognitive load is more important than generation, thus worked examples will lead to the highest learning outcomes.
2. Questionnaire: We predicted that worked examples would lead to the lowest cognitive load ratings, and invention would lead to the highest, with completion problems in the middle. We hypothesized that perceived knowledge gaps and interest would be equal or higher in the invention condition compared to the other two conditions (Glogger-Frey et al., 2015).

3.1. Method

3.1.1. Participants

Undergraduate students ($N = 123$; age $M = 19.02$, $SD = 2.04$; 63.6% female) participated for research credit in an introductory psychology course. As reported above, previous experimental studies manipulating guidance during exploratory learning have shown medium to large effect sizes. A G*Power analysis for ANCOVA ($\alpha = 0.05$, power = .95, df = 2, groups = 3, covariates = 1; Faul, Erdfelder, Buchner, & Lang, 2009) showed that a sample size of 129 would be sufficient to achieve $\eta_p^2 = 0.11$ (f = 0.35; medium effect = 0.25, large = 0.40). Students were randomly assigned to one of three conditions: Invention ($n = 40$), completion problems ($n = 42$), or worked examples ($n = 41$). Four additional students were excluded from analyses for failure to complete the posttest.

(a) Invention Condition

Year	Tea growers: Thourbo (Antioxidants per mg)	Dareen	Ging
2010	--	14	11
2011	10	19	12
2012	16	14	14
2013	16	17	21
2014	20	17	18
2015	13	9	14

Come up with a formula for consistency to show which tea grower has the most consistent levels of antioxidants. Show your proposed formulas and calculations on the next page. You may also list steps for how you would calculate consistency. Circle the tea grower that you decide is most consistent.

(b) Completion Problems Condition

Step 2: Calculate the Dispersion

$(10 - 15)^2 = -5^2 = 25$

$(16 - 15)^2 = 1^2 = 1$

$(16 - \square)^2 = 1^2 = 1$

$(\square - 15)^2 = 5^2 = \square$

$(\square - \square)^2 = -2^2 = \square$

Step 2: Next, we need to find the *dispersion*: how far apart scores are from the mean. We can find out the distance of a score from the mean by subtracting the mean from that score.

If you stopped there, some of the differences would result in a negative value, and some would result in a positive value. To eliminate any negative values, we square each difference score.

Finally, we add up the differences.

(c) Worked Examples Condition

Step 2: Calculate the Dispersion

$(10 - 15)^2 = -5^2 = 25$

$(16 - 15)^2 = 1^2 = 1$

$(16 - 15)^2 = 1^2 = 1$

$(20 - 15)^2 = 5^2 = 25$

$(13 - 15)^2 = -2^2 = 4$

Step 2: Next, we need to find the *dispersion*: how far apart scores are from the mean. We can find out the distance of a score from the mean by subtracting the mean from that score.

If you stopped there, some of the differences would result in a negative value, and some would result in a positive value. To eliminate any negative values, we square each difference score.

Finally, we add up the differences.

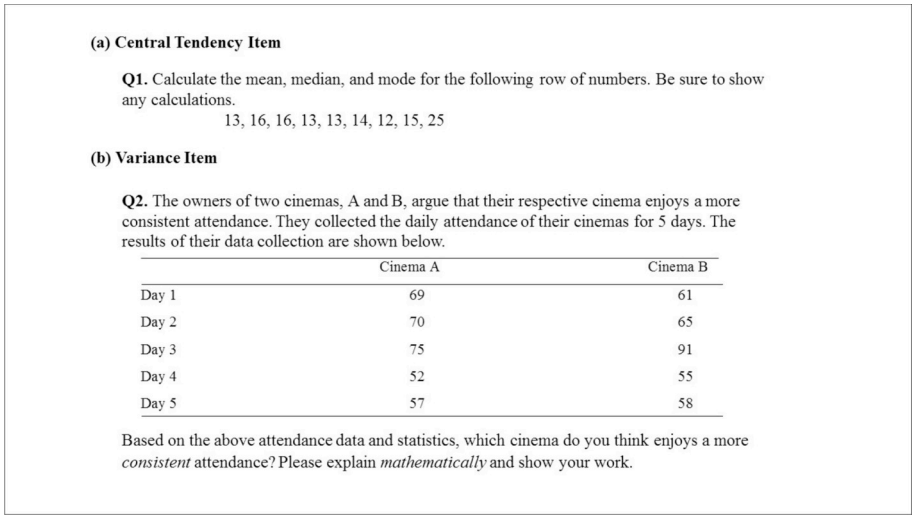
Fig. 1. Examples of the three activities used during exploration in Experiment 1.

3.1.2. Materials

The activity, instruction, and posttest are included in Appendices A – C.

Pretest. The pretest included two items (Fig. 2). A central tendency problem asked students to find the mean, median, and mode of an array of nine numbers (1 point each; Paas, 1992). A variance problem (adapted from Kapur, 2012) provided a table of attendance data for two cinemas over five days, and asked students to determine mathematically which of the two cinemas enjoys the most consistent attendance (4-points possible; Appendix D).

Problem-Solving Activity. The problem-solving activity (adapted from Wiedmann, Leach, Rummel, & Wiley, 2015) asked students 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 six years. Students in the *invention condition* were instructed to invent a formula, or list the steps used, to calculate consistency for each tea grower (Roll, Aleven, & Koedinger, 2009). As shown in Fig. 1, students in the *worked examples condition* received the same problem with standard deviation worked out for each tea grower, with brief explanations for the calculations.



(a) Central Tendency Item

Q1. Calculate the mean, median, and mode for the following row of numbers. Be sure to show any calculations.

13, 16, 16, 13, 13, 14, 12, 15, 25

(b) Variance Item

Q2. The owners of two cinemas, A and B, argue that their respective cinema enjoys a more consistent attendance. They collected the daily attendance of their cinemas for 5 days. The results of their data collection are shown below.

	Cinema A	Cinema B
Day 1	69	61
Day 2	70	65
Day 3	75	91
Day 4	52	55
Day 5	57	58

Based on the above attendance data and statistics, which cinema do you think enjoys a more *consistent* attendance? Please explain *mathematically* and show your work.

Fig. 2. Pretest items used in Experiments 1 and 3.

Students were instructed to study the calculations. The *completion problems condition* was the same as the worked example condition, with some empty boxes in place of parts of the calculations. Empty boxes were chosen to ensure encoding of various parts of the calculations (e.g., dispersion of scores from the mean, sum of squares, dividing by *n*, taking square root), and the number of boxes increased per tea grower. Students were instructed to complete the calculations.

Questionnaire. Cognitive load was measured with the Mental Effort Rating Scale (Paas, 1992). Students responded to the prompt ("In solving or studying the previous problem I invested …") on a scale from 1 (very, very low mental effort) to 9 (very, very high mental effort). Interest and enjoyment were measured with three items (McDonald's ω = 0.88) adapted from Ryan's (1982) Intrinsic Motivation Inventory. Perceived knowledge gaps were measured with four items (McDonald's ω = 0.89) adapted from Flynn and Goldsmith (1999). Items from both scales were intermixed and rated on a 5-point Likert scale (1 = strongly disagree; 5 = strongly agree; Table 1).

Instruction. Direct instruction was provided as a text passage (adapted from Wiedmann et al., 2015). Students were given a worked example, in addition to conceptual explanations and definitions. Students were told that engineers were interested in comparing which trampoline (A or B) has the most consistent levels of bounciness. A table displayed data for inches of rebound for trampoline A, followed by the formula and instructions to calculate standard deviation. Text boxes explained concepts and calculations. Then a table displayed inches of rebound for Trampoline B, followed by three questions to help students practice. The first two questions asked students for the standard deviation of Trampoline B, and what this value meant. Finally, students were instructed to determine which of the two trampolines has the most consistent bounciness.

Table 1
Interest and enjoyment and knowledge gap items.

Interest and enjoyment
1 I found this learning activity interesting.
2 I enjoyed this learning activity.
3 This learning activity was boring. (reverse coded)

Knowledge gaps
1 I do not feel very knowledgeable about calculating consistency.
2 When it comes to calculating consistency, I really don't know a lot.
3 Compared to most other people, I know less about calculating consistency.
4 I know pretty much about calculating consistency. (reverse coded)

Posttest. Three posttest items were drawn from Wiedmann et al. (2012; see also Kapur, 2012). One was from a psychological statistics exam used at the students' university, and was similar to Schwartz and Martin's (2004) symbolic insight problems.

A *procedural fluency item* asked students to determine in which of two months an ice hockey tournament should be held. Students were given a dataset including daily temperatures for each month and instructed to select the month with the most consistent temperatures. Students were asked to explain their decision mathematically.

An item measuring *conceptual understanding* included the same dataset, and students were told that one of the values was incorrect. The correct value was provided, and students were asked to determine whether or not their choice from the previous item should change based on this new information. They were further asked to explain why this mistake mattered or not. Another conceptual understanding item listed two components of the standard deviation formula (the numerator: $(x–M)$, $(\ )^2$, and Σ; and the square root √). Students were asked to explain how each component contributes to the concept of standard deviation.

Finally, a *transfer* item instructed students to determine which of two students, the top physics student or the top chemistry student, deserves the best science student award. A table displayed the scores of the top physics and chemistry students for five years. The top students from the current year both had higher scores than students in the previous years. Students were asked which of the two current students deserves the award more, and to explain their decision mathematically. This problem is solved by calculating the standard deviation for both physics and chemistry students, and then determining which of the two current students is more standard deviations away from their disciplines' mean score. This item required students to build upon their conceptualization of standard deviation to develop standardized scores (Kapur, 2012).

All items were scored on a four-point scale (Appendix D). Twenty percent of the items were scored by a second observer; interrater reliability was high (r = 0.90).

3.1.3. Procedure

Students completed the study individually in sessions of up to fifteen in a reserved classroom. After providing consent, students were instructed that the purpose of the study was to see how people learn from various activities, and that they would learn about calculating consistency in statistics. Students then completed an individual differences questionnaire and pretest (8 min). The 35-item questionnaire was administered as part of a larger study and will not be discussed further.

Afterwards, students worked on the problem-solving activity (15 min). Activities were administered on paper and interleaved by condition, and students were randomly assigned to condition based on the activity they received. Following, students were given the questionnaire followed by written instruction (15 min). Then students completed the posttest (30 min) and were debriefed.

3.2. Results and discussion

3.2.1. Preliminary analyses

Pretest items were examined as a function of activity type, revealing no effect on the central tendency item (invention: $M = 2.04$ out of 3, $SD = 0.92$; completion problems: $M = 2.31$, $SD = 0.87$; worked examples: $M = 2.50$, $SD = 0.74$), $F(2,120) = 1.52$, $p = .224$. However, activity type had a significant effect on the variance item (invention: $M = 1.10$ out of 4, $SD = 0.65$; completion problems: $M = 1.52$, $SD = 0.67$; worked examples: $M = 1.02$, $SD = 0.80$), $F(2,120) = 5.75$, $p = .004$. Despite random assignment, prior knowledge of variance was unequal across conditions. Thus, this variable was used as a covariate in all subsequent analyses. The pattern of results reported below remains statistically significant even without this covariate in analyses.

3.2.2. Learning outcomes

The distributions of posttest scores are shown in Fig. 3. Posttest scores were examined using a 3 (activity type: invention, completion problems, worked examples) × 3 (posttest subscale: procedural, conceptual, transfer) repeated measures ANCOVA, with activity type as a between-subjects factor, posttest subscale as a within-subjects factor, and scores on the pretest variance item as a covariate. We predicted that students who explored in the worked examples and completion problems conditions would demonstrate higher posttest scores than those who explored in the invention condition. We tested two possible predictions for completion problems compared to worked examples: higher scores for completion problems due to generating additional problem information, or lower scores if generation is less important.

The assumption of sphericity was violated, $p = .017$. Therefore, the lower-bound statistic was used. A non-significant main effect of activity type was found, $F(2,119) = 3.01$, $p = .053$, $\eta_p^2 = 0.05$ (Table 2, Fig. 4). Planned comparisons revealed that, as predicted, studying worked examples led to significantly higher posttest scores than invention, $t(119) = 2.37$, $p = .019$, $d = 0.54$. This result replicates the effects found by Glogger-Frey et al. (2015) using different subject matter. However, completion problems did not improve posttest scores compared to invention, $t(119) = 1.68$, $p = .098$, $d = 0.36$, or worked examples, $t(119) = 0.73$, $p = .469$, $d = 0.15$. Thus, completion problems showed a middling effect.

A main effect of posttest subscale was found, $F(1,119) = 16.05$, $p < .001$, $\eta_p^2 = 0.12$. Post-hoc comparisons with Bonferroni correction ($\alpha = 0.016$) revealed that students scored higher on procedural ($M = 3.30$, $SD = 0.94$) compared with conceptual ($M = 2.23$, $SD = 1.10$), $t(119) = 11.89$, $p < .001$, $d = 1.06$, and transfer ($M = 2.41$, $SD = 1.03$) subscales, $t(119) = 9.84$, $p < .001$, $d = 0.90$. Conceptual and transfer subscales did not differ significantly, $t(119) = -1.68$, $p = .094$, $d = 0.15$. There was no interaction between activity type and posttest subscale, $F < 1$, indicating that the effects of activity type were similar across the subscales.

3.2.3. Questionnaire

Questionnaire data were examined with ANCOVAs (Table 2). A significant effect of activity type was found for cognitive load, $F(2,118) = 3.20$, $p = .045$, $\eta_p^2 = 0.05$. Planned comparisons demonstrated that, as hypothesized, worked examples led to less cognitive load than invention, $t(118) = -2.39$, $p = .019$, $d = 0.60$. Also as predicted, completion problems did not lead to less cognitive load than either worked examples, $t(118) = 0.48$, $p = .630$, $d = 0.09$, or invention, $t(118) = -1.87$, $p = .063$, $d = 0.45$. Thus, cognitive load for completion problems was between that of the other two conditions.

The finding that completion problems did not significantly impact posttest scores or cognitive load ratings compared to invention conflicts with documented benefits of generation on memory and learning (Slamecka & Graf, 1978). These findings do align with others demonstrating no benefit of partial guidance during exploratory learning (cf. Alfieri et al., 2011; Likourezos & Kalyuga, 2016; Loibl & Rummel, 2014b; but see; Borek, McLaren, Karabinos, & Yaron, 2009). A key difference between studies is the form of guidance used. Rather than using worked examples, Borek et al. provided hints and feedback during a learning activity. Thus, results may differ depending on the type of guidance used.

An effect of condition was found for perceived knowledge gaps, $F(2,117) = 13.83$, $p < .001$, $\eta_p^2 = 0.19$. As predicted, completion problems, $t(117) = -4.65$, $p < .001$, $d = 1.16$, and worked examples, $t(117) = -4.36$, $p < .001$, $d = 1.21$, led to significantly lower knowledge gaps than invention. The finding that inventing led to greater perceived knowledge gaps than worked examples, but did not lead to superior learning, also replicates Glogger-Frey et al. (2015).

We had hypothesized that interest and enjoyment would be either higher in the invention condition or equal across conditions. Interest and enjoyment did not differ as a function of condition, $F(2,116) = 1.62$, $p = .203$, $\eta_p^2 = 0.02$.

3.2.4. Conclusion

These findings suggest that worked examples maintain the benefits of exploratory learning while reducing cognitive load. Previous studies have only examined guided alternatives to invention within an exclusively exploratory learning context (Glogger-Frey et al., 2015; Likourezos & Kalyuga, 2016), or have only included direct instruction conditions in which instruction is followed by unguided practice problems (e.g., Loibl & Rummel, 2014b). These findings cannot speak to whether worked examples actually enable exploratory learning or whether they simply function as another method of providing instruction first. We provide this comparison in Experiment 2.

4. Experiment 2

Experiment 2 compared the impact of guidance during problem solving in both an exploratory learning and traditional instruction setting, using a 2 (activity type: invention, worked examples) × 2 (order of instruction: explore-first, instruct-first) between-subjects factorial design. Because completion problems did not show learning differences compared to the other conditions in Experiment 1, this condition was not included. We also attempted to replicate and extend the findings from Experiment 1 to an undergraduate psychological statistics course, in order to gauge the ecological validity of the findings. The materials were the same as in Experiment 1, except that we excluded the pretest due to concerns that the pretest may achieve instructional goals inherent to minimally guided exploration (e.g., Kapur, 2016).

We tested the following hypotheses:

1. Learning Outcomes: We predicted an overall benefit of exploratory learning, whereby students in the explore-first conditions would outperform those in the instruct-first conditions on the posttest. Based on Experiment 1, we also hypothesized that those who explored using worked examples would outperform those who explored using invention.
2. Questionnaire: We hypothesized that cognitive load and perceived knowledge gaps would be highest for those who invented prior to instruction, compared to other conditions.

4.1. Method

4.1.1. Participants

Participants were 190 undergraduate students (Age $M = 20.67$,

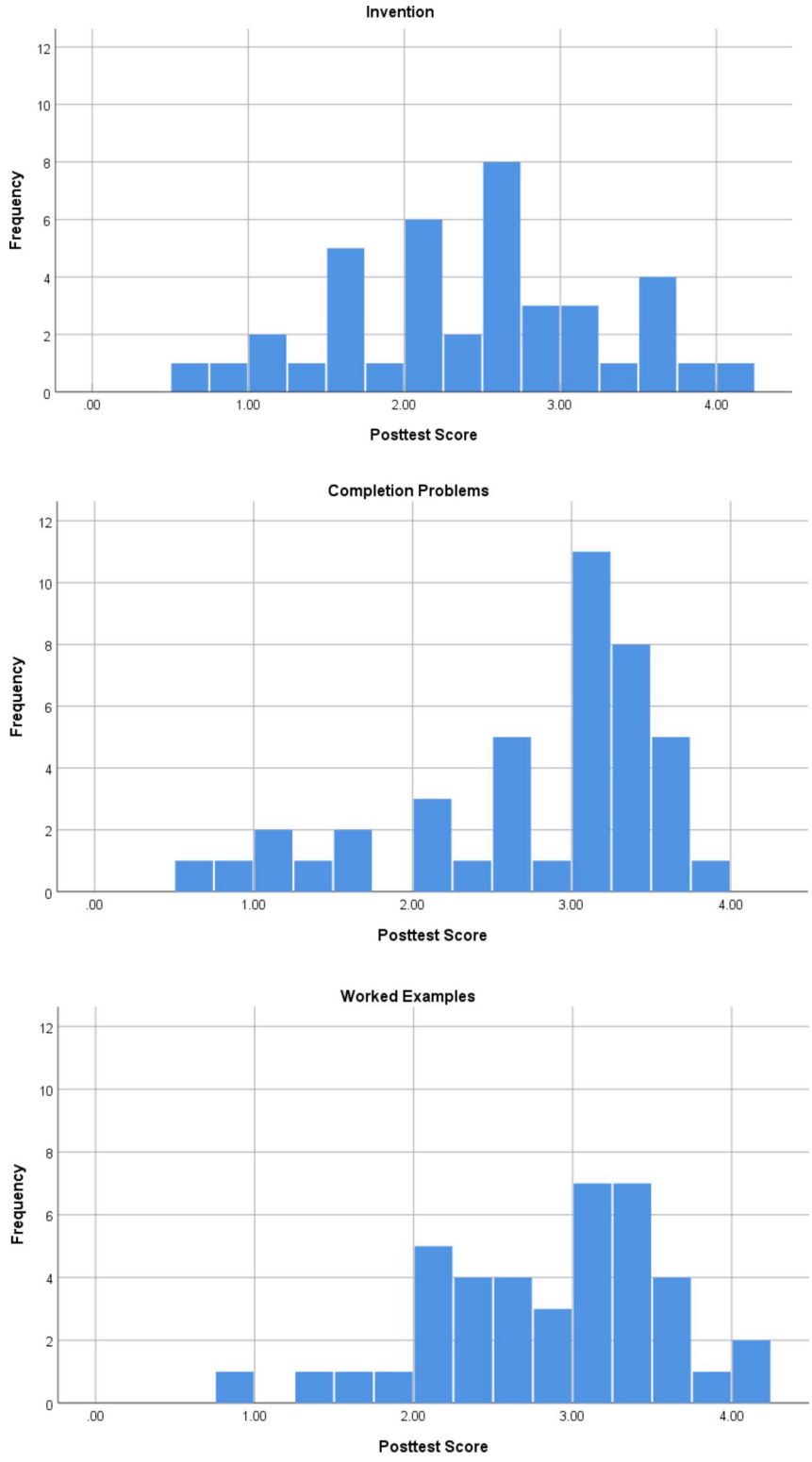


Fig. 3. Distribution of posttest scores for each condition in Experiment 1.

SD = 4.33; 72.9% female) enrolled in three sections of a psychological statistics course across two semesters, with two different instructors. Students were randomly assigned to one of four conditions: Explore-first/worked examples (n = 46), explore-first/invention (n = 48), instruct-first/worked examples (n = 47), or instruct-first/invention (n = 49). A G*Power analysis for ANCOVA (α = 0.05, power = .95, df = 3, groups = 4, covariates = 1; Faul et al., 2009) showed that a sample size of 195 would be enough to achieve a medium effect (f = 0.30). Additional students were excluded from analyses for failure to provide consent (n = 3), failure to complete the posttest (n = 11), absence or inability to link their posttest to their first session materials (e.g., no name on the paper; n = 24), or for having participated in Experiment 1 (n = 5).

4.1.2. Materials

The materials used in Experiment 2 were identical to Experiment 1,

Table 2
Means (standard deviations in parentheses) of posttest and questionnaire data as a function of activity type (Experiment 1).

	Invention	Completion Problems	Worked Examples
Posttest Score	2.40 (0.85)	2.71 (0.85)	2.83 (0.73)
Cognitive Load	5.70 (0.99)	4.98 (2.02)	4.80 (1.86)
Knowledge Gaps	3.90 (0.80)	2.79 (1.09)	2.85 (0.93)
Interest and Enjoyment	3.13 (0.91)	3.43 (0.84)	3.20 (0.95)

Note: Posttest scores were out of 4 possible points. Cognitive load was measured with a 9-point Likert scale. Knowledge gaps and interest and enjoyment were measured with a 5-point Likert scale.

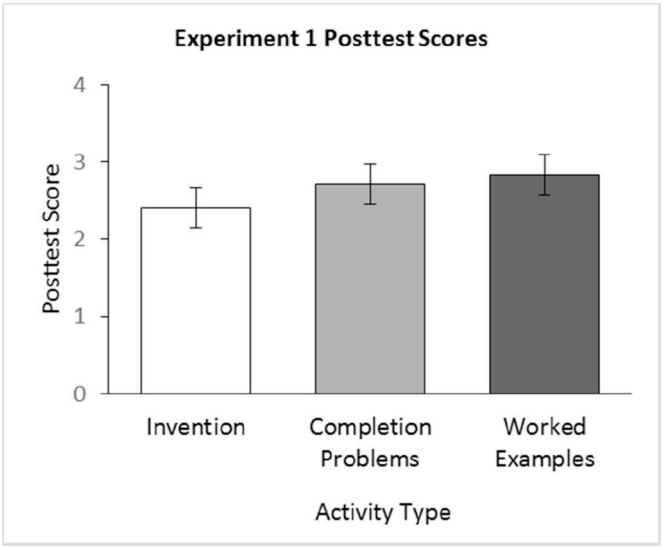


Fig. 4. Experiment 1 mean posttest scores as a function of activity type. Error bars represent 95% confidence intervals.

aside from three changes: (1) The pretest was cut from the procedure; (2) A prompt in the worked example asked students if they agreed with the chosen tea grower; (3) In the instruct-first condition activities, students were asked to use what they had just learned about standard deviation to help the managers determine which tea grower produces the most consistent levels of antioxidants, rather than to come up with this formula; (4) The consent form, problem-solving activity, questionnaire, and instruction were combined into one set of paper materials, with signals to stop and wait for instruction at the end of each section.

The invention and worked examples conditions were the same as in Experiment 1. These activities were provided either before instruction (explore-first conditions) or after instruction (instruct-first conditions). The questionnaire always followed the activity. A second rater scored 20% of the posttests (interrater reliability: $r = 0.90$).

4.1.3. Procedure

Students completed the study across two course lab sessions that were 1–2 weeks apart, occurring prior to lectures covering standard deviation and variance. Students were told that the activities would help them learn about concepts relevant to the course, and to try their best, but their performance on the activities would not affect their grade. The first session included the problem solving activity, questionnaire, and direct instruction. The second session included the posttest. Students were provided with calculators.

Students were randomly assigned to condition based on the materials they received, which were interleaved by condition. After reviewing the consent form, students completed the first section (problem-solving activity/questionnaire or direct instruction, depending on condition; 15 min). Students then completed the second section (problem-solving activity/questionnaire or direct instruction, depending on condition; 15 min). In the second session, students completed the posttest (30 min) and a demographics questionnaire, and were debriefed.

4.2. Results and discussion

Although the course instructors did not administer the learning materials, we controlled for possible differences based on instructor in all analyses.

4.2.1. Learning outcomes

Distributions of posttest scores are shown in Fig. 5. Posttest performance was examined with a 3 (posttest subscale: procedural, conceptual, transfer) × 2 (order of instruction: explore-first, instruct-first) × 2 (activity type: invention, worked examples) ANCOVA, with posttest subscale as a within-subjects factor and order of instruction and activity type as between-subjects factors. The assumption of sphericity was violated, $p < .001$, so the lower-bound statistic was used.

As in Experiment 1, a significant effect of posttest subscale was found $F(1,185) = 32.74$, $p < .001$, $\eta_p^2 = 0.14$. Students scored higher on procedural ($M = 3.03$, $SD = 1.12$) than conceptual ($M = 2.21$, $SD = 1.21$), $t(185) = 13.83$, $p < .001$, $d = 1.01$, and transfer subscales ($M = 2.53$, $SD = 1.19$), $t(185) = 6.52$, $p < .001$, $d = 0.48$. Transfer scores were higher than conceptual scores, $t(185) = 4.47$, $p < .001$, $d = 0.32$.

We predicted an overall benefit of exploratory learning over instruct-first conditions. Supporting this hypothesis, a main effect of order of instruction was found. Those in the explore-first condition ($M = 2.75$, $SD = 1.04$) outperformed their instruct-first ($M = 2.44$, $SD = 1.0$) counterparts, $F(1,185) = 4.29$, $p = .040$, $\eta_p^2 = 0.02$ (Fig. 6). In contrast to our hypothesis, there was no main effect of activity type (invention: $M = 2.57$, $SD = 1.01$; worked examples: $M = 2.31$, $SD = 0.98$) or interaction, $Fs < 1$. For those in the explore-first conditions, a planned comparison revealed no difference between the worked examples or invention conditions, $t(185) = 0.15$, $p = .878$, $d = 0.03$ (Table 3). Similarly, in the instruct-first conditions, there was no difference between the worked examples and invention conditions, $t(185) = 1.22$, $p = .222$, $d = 0.55$.

4.2.2. Questionnaire

An order of instruction × activity type ANCOVA revealed a significant main effect of activity type on cognitive load (Table 3). As predicted, those in the worked examples conditions reported lower cognitive load ($M = 4.91$, $SD = 1.64$) than those in the invention conditions ($M = 5.57$, $SD = 1.49$), $F(1,175) = 7.75$, $p = .006$, $\eta_p^2 = 0.04$. Order of instruction did not affect cognitive load, $F(1,175) = 1.07$, $p = .302$, $\eta_p^2 = 0.01$. There was no interaction, $F(1,175) = 2.82$, $p = .095$, $\eta_p^2 = 0.02$. Planned comparisons showed that, in the explore-first conditions, those who studied worked examples reported lower cognitive load than those who invented, $t(175) = -3.24$, $p = .001$, $d = 0.72$. In the instruct-first conditions, cognitive load was comparable between activity types, $t(175) = -0.76$, $p = .449$, $d = 0.15$.

For perceived knowledge gaps, an ANCOVA revealed a significant main effect of order of instruction. As predicted, students in the explore-first conditions ($M = 3.36$, $SD = 1.05$) reported greater knowledge gaps than those in the instruct-first conditions ($M = 2.76$, $SD = 0.96$), $F(1,175) = 16.99$, $p < .001$, $\eta_p^2 = 0.09$. There was also a main effect of activity. Students who invented ($M = 3.33$, $SD = 1.06$) reported greater knowledge gaps than those who studied worked examples ($M = 2.83$, $SD = 1.05$), $F(1,175) = 9.07$, $p = .003$, $\eta_p^2 = 0.05$. These effects were qualified by a significant interaction, $F(1,175) = 20.46$, $p < .001$, $\eta_p^2 = 0.11$ (Table 3). As predicted, students who explored reported greater knowledge gaps in the invention condition than in the worked examples condition, $t(175) = 5.49$, $p < .001$, $d = 1.16$. Students who

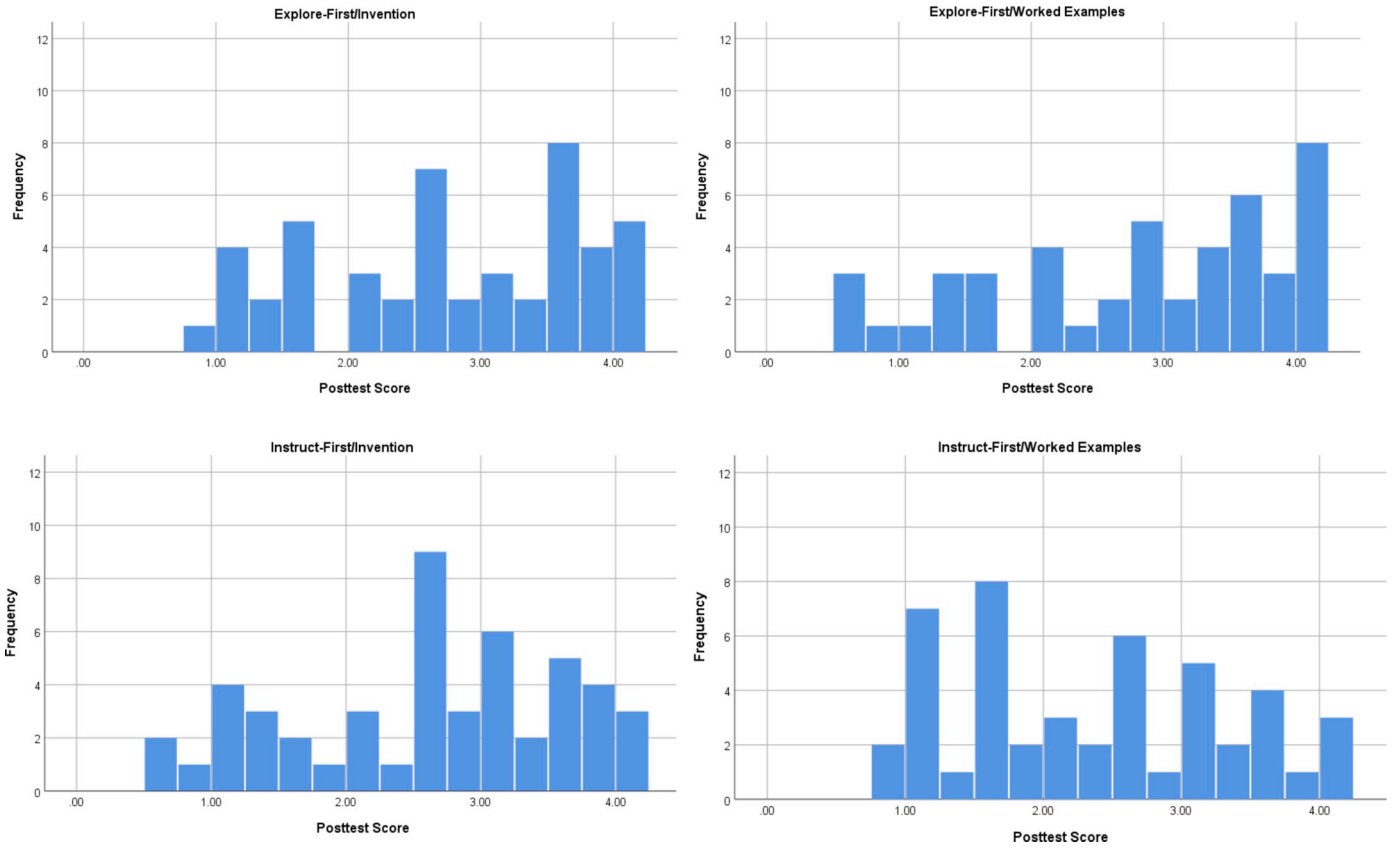


Fig. 5. Distribution of posttest scores for each condition in Experiment 2.

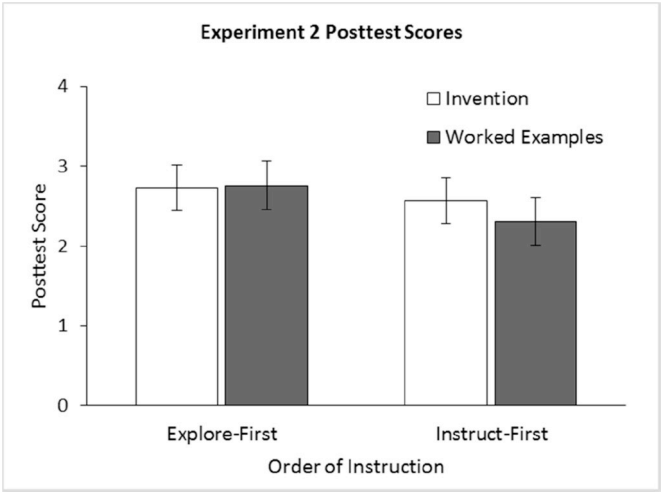


Fig. 6. Experiment 2 mean posttest scores as a function of order of instruction and activity type. Error bars represent 95% confidence intervals.

received instruction first reported similar knowledge gaps across both activity types, $t(175) = 1.05$, $p = .351$, $d = 0.23$.

For interest and enjoyment, an ANCOVA revealed no significant main effects of order of instruction, $F(1,175) = 3.38$, $p = .068$, $\eta_p^2 = 0.02$, activity type, $F < 1$, or interaction, $F(1,175) = 1.25$, $p = .265$, $\eta_p^2 = 0.01$ (Table 3).

4.2.3. Conclusion

Thus, on average, students in the explore-first conditions outperformed students in the instruct-first conditions on the posttest, regardless of the guidance they received. This finding is consistent with others demonstrating the benefits of exploratory learning prior to instruction, adding to this literature a tightly-controlled study in a psychological statistics classroom setting.

However, this experiment did not replicate the results of activity type in Experiment 1. In the explore-first conditions, students who studied worked examples reported lower cognitive load and knowledge gaps, yet demonstrated comparable learning to students who invented. Two differences between experiments might account for these discrepant results. First, the sample in Experiment 1 was recruited for a

Table 3
Means (standard deviations in parentheses) of posttest scores and questionnaire data as a function of order of instruction and activity type (Experiment 2).

	Explore-First		Instruct-First	
	Invention	Worked Examples	Invention	Worked Examples
Posttest Score	2.73 (0.99)	2.76 (1.10)	2.57 (1.00)	2.31 (0.98)
Cognitive Load	5.65 (1.31)	4.59 (1.61)	5.49 (1.69)	5.24 (1.64)
Knowledge Gaps	3.88 (0.84)	2.82 (0.97)	2.66 (0.92)	2.88 (0.99)
Interest and Enjoyment	3.19 (0.82)	3.34 (0.74)	3.56 (0.94)	3.44 (0.83)

Note: Posttest scores were out of 4 possible points. Cognitive load was measured with a 9-point Likert scale. Knowledge gaps and interest and enjoyment were measured with a 5-point Likert scale.

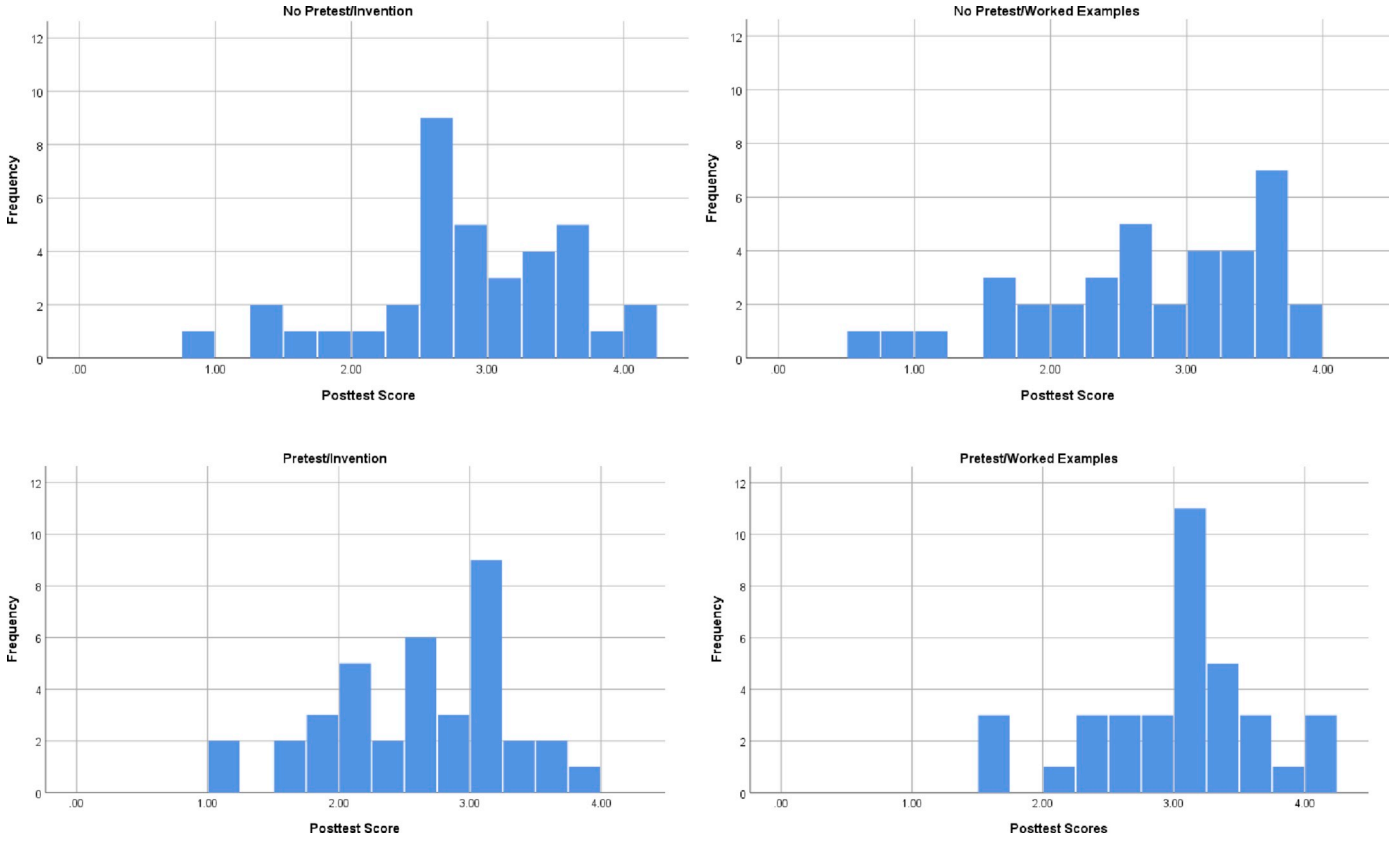


Fig. 7. Distribution of posttest scores for each condition in Experiment 3.

laboratory-based experiment from an introductory psychology course typically taken by first-year students. The sample in Experiment 2 included students in a psychological statistics course typically taken in the second or third year. Students in Experiment 2 may have had more prior knowledge and motivation to learn the material.

Experiment 1 also included a pretest, whereas Experiment 2 did not. Pretests are commonly used to assess and control for prior knowledge. However, a pretest can function as a learning tool, if the pretest targets the to-be-learned concept (e.g., Richland et al., 2009). The pretest included an item similar to the learning activity, asking students to determine which of two cinemas enjoyed the most consistent attendance. Students may have engaged in aspects of exploration while completing the pretest. Thus, worked examples may only facilitate exploration in combination with a pretest. This possibility was tested in Experiment 3.

5. Experiment 3

Experiment 3 was designed to reconcile the contradictory findings of Experiments 1 and 2 by testing the impact of a pretest, keeping the sample constant across conditions. We conducted a laboratory-based experiment using a 2 (activity type: invention, worked examples) × 2 (pretest condition: pretest, no-pretest) between-subjects factorial design. All students completed exploratory learning activities, either inventing or exploring worked examples prior to instruction, with or without a pretest.

We tested the following hypotheses:

1. Learning Outcomes: We hypothesized that worked examples would enhance posttest scores relative to invention only when a pretest was administered.
2. Questionnaires: We hypothesized that worked examples would lead to lower cognitive load ratings, lower knowledge gaps, and comparable interest and enjoyment, compared to invention. We also explored how these mechanisms were impacted by the pretest.

5.1. Method

5.1.1. Participants

Participants were 147 undergraduate students (Age $M = 19.47$, $SD = 4.46$; 53% female) recruited for research credit in an introductory psychology course. Students were randomly assigned to one of four conditions: No-pretest/invention ($n = 37$), no-pretest/worked examples ($n = 37$), pretest/invention ($n = 37$), or pretest/worked examples ($n = 36$). A G*Power analysis ($\alpha = 0.05$, power = .95, df = 3, groups = 4; Faul et al., 2009) showed that a sample of 145 would be enough to achieve a medium effect ($f = 0.35$). Additional students were excluded for falling asleep during the experiment ($n = 1$) and failure to complete the posttest ($n = 5$).

5.1.2. Materials

The activities, instruction, and questionnaires were identical to Experiment 1, with one exception. Due to an error, the prompt for the invention activity was the same as that for the instruct-first condition in Experiment 2. Across all experiments, the activity cover story stated that the managers "want a formula for calculating the consistency of antioxidant levels for each tea grower" and that "this formula should apply to all tea growers and help provide a fair comparison." However, in the previous studies' explore-first conditions, students were asked to come up with a formula for consistency, and to show their proposed formulas or steps. In Experiment 3, students were instructed to use what they just learned about standard deviation to help determine which tea grower produces the most consistent levels of antioxidants, and to show their calculations.

Additionally, the questionnaire was administered twice in the

pretest conditions, once following the pretest, and once following the activity.

5.1.3. Procedure

Sessions included groups of up to 15 in a reserved classroom. Due to differences in timing, each session included either the pretest conditions or the no-pretest conditions. Following consent, students in the pretest conditions completed the pretest and post-pretest questionnaire (6 min). Then, all students began the exploratory learning activity. Students were randomly assigned to activity condition based on the materials they received (15 min). Following, students completed the post-activity questionnaire and direct instruction (15 min). Finally, students completed the posttest (30 min) and a demographics questionnaire, then were debriefed. Posttests were scored by two independent scorers (interrater reliability: $r = 0.93$).

5.2. Results and discussion

5.2.1. Preliminary analyses

In the pretest conditions, prior knowledge was equal across the invention (central tendency: $M = 2.14$, $SD = 1.03$; variance: $M = 1.46$, $SD = 0.61$) and worked examples conditions (central tendency: $M = 2.41$, $SD = 0.77$; variance: $M = 1.50$, $SD = 0.56$), $ts(71) = -1.31$ and -0.297, $ps = .192$ and .768, $ds = -0.30$ and -0.07, respectively.

5.2.2. Learning outcomes

The distributions of posttest scores are shown in Fig. 7. Scores were analyzed with a 3 (subscale: procedural, conceptual, transfer) × 2 (activity type: invention, worked examples) × 2 (pretest condition: pretest, no-pretest) mixed-factorial ANOVA, with subscale as a within-subjects factor and activity type and pretest condition as between-subjects factors. The assumption of sphericity was violated, $p < .001$, so the lower-bound statistic was used.

As in Experiments 1 and 2, a main effect of posttest subscale was found, $F(2,143) = 97.68$, $p < .001$, $\eta_p^2 = 0.40$. Students scored higher on procedural ($M = 3.39$, $SD = 0.70$) than conceptual ($M = 2.23$, $SD = 0.95$), $t(143) = 17.31$, $p < .001$, $d = 1.41$, and transfer ($M = 2.63$, $SD = 1.17$) subscales, $t(143) = 8.36$, $p < .001$, $d = 0.68$. Transfer scores were higher than conceptual, $t(143) = 4.29$, $p < .001$, $d = 0.36$. No main effects of pretest, $F(1,143) = 1.50$, $p = .222$, $\eta_p^2 = 0.01$, or activity type, $Fs < 1$, were found. No interactions including subscale were significant, $Fs \leq 2.76$, $ps > .098$. Thus, any effect of condition occurred similarly across subscales.

There was a significant activity type × pretest condition interaction, $F(1,143) = 4.28$, $p = .040$, $\eta_p^2 = 0.03$ (Fig. 8; Table 5). We hypothesized that worked examples would enhance exploratory learning compared to invention only when used in combination with a pretest. Results supported this prediction. For the pretest conditions, planned comparisons revealed that students who explored worked examples had higher posttest scores than those who invented, $t(143) = 2.33$, $p = .022$, $d = 0.62$. In the no-pretest conditions, worked examples did not lead to better posttest performance than inventing, $t(143) = -0.60$, $p = .550$, $d = 0.15$.

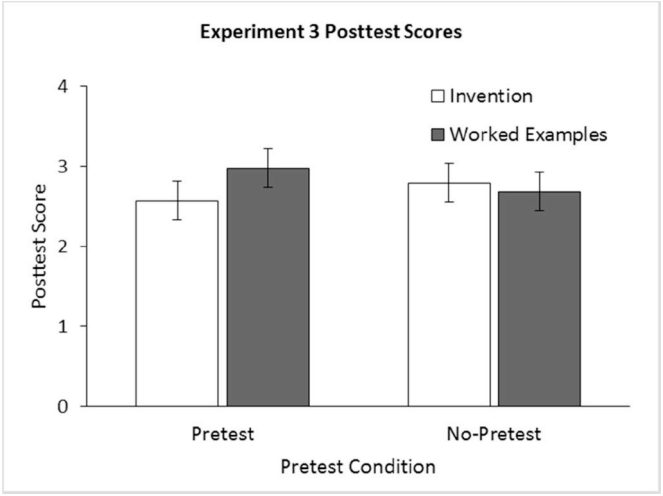


Fig. 8. Experiment 3 mean posttest scores as a function of pretest condition and activity type. Error bars represent 95% confidence intervals.

As shown in Fig. 8 and Table 5, the two pretest conditions resulted in the highest and lowest posttest averages, with the no-pretest conditions in the middle. However, scores for the pretest/worked examples condition were not significantly higher than those of the no-pretest conditions, $ts \leq 1.67$, $ps \geq = .096$, $ds \leq 0.39$. Moreover, scores for the pretest-invention condition were not significantly lower than the two no-pretest conditions, $ts \leq -0.65$, $ps \geq .214$, $ds \leq 0.31$. Thus, although the pretest had a divergent effect on activity type, these results cannot speak to whether the pretest improves exploring with worked examples or decreases exploring with invention (or both).

5.2.3. Post-pretest questionnaire

We first analyzed questionnaire data for students in the two pretest conditions, to examine changes from post-pretest to post-activity (Table 4). Each questionnaire was examined using a 2 (questionnaire timing: post-pretest, post-activity) × 2 (activity type: invention, worked examples) ANOVA, with questionnaire timing as a within-subjects factor and activity type as a between-subjects factor.

For cognitive load ratings, no main effects were found for questionnaire timing, $F(1,63) = 1.47$, $p = .231$, $\eta_p^2 = 0.02$, or activity, $F < 1$. However, there was a significant interaction, $F(1,63) = 12.26$, $p = .001$, $\eta_p^2 = 0.16$. For students who invented, cognitive load increased from post-pretest to post-activity, $t(63) = -3.46$, $p = .001$, $d = 0.52$. For students who explored worked examples, cognitive load remained the same, $t(63) = 1.56$, $p = .124$, $d = 0.27$. Although cognitive load increased for the pretest/invention condition, cognitive load was not significantly higher than in the pretest/worked examples condition either post-pretest, $t(63) = 0.56$, $p = .578$, $d = 0.31$, or post-activity, $t(63) = 1.79$, $p = .078$, $d = 0.45$.

For knowledge gaps, there was no main effect of questionnaire timing, $F(1,61) = 3.51$, $p = .066$, $\eta_p^2 = 0.05$. A main effect of activity was found, $F(1,61) = 5.61$, $p = .021$, $\eta_p^2 = 0.08$. Students in the pretest/invention condition ($M = 3.16$, $SD = 0.75$) perceived greater knowledge gaps overall than students in the pretest/worked examples condition ($M = 2.75$, $SD = 0.62$). These effects were qualified by a significant interaction, $F(1,61) = 16.43$, $p < .001$, $\eta_p^2 = 0.21$. In the pretest/worked examples condition, knowledge gaps dropped from post-pretest to post-activity, $t(61) = 3.98$, $p < .001$, $d = 0.59$. In the pretest/invention condition, knowledge gaps remained constant post-pretest to post-activity, $t(61) = -1.63$, $p = .107$, $d = 0.39$. At post-pretest, knowledge gaps were the same for both pretest conditions, $t(61) = 0.49$, $p = .629$, $d = 0.12$. Following the activity, knowledge gaps were greater in the pretest/invention condition than in the pretest/worked examples condition, $t(61) = 4.57$, $p < .001$, $d = 1.16$. Thus, the pretest instantiated knowledge gaps in both pretest conditions. However, these knowledge gaps decreased in the worked examples condition relative to the invention condition.

A main effect of questionnaire timing was found for interest and enjoyment, $F(1,62) = 4.70$, $p = .034$, $\eta_p^2 = 0.07$. Ratings were higher following the activity ($M = 3.44$, $SD = 0.82$) than following the pretest ($M = 3.29$; $SD = 0.78$). There was no main effect of activity type or interaction, $Fs < 1$.

5.2.4. Post-activity questionnaire

We next examined post-activity questionnaire ratings across all conditions (Table 5). For cognitive load, an activity type × pretest condition ANOVA revealed no main effects of activity type, F

Table 4
Means (standard deviations in parentheses) for post-pretest and post-activity questionnaires in both pretest conditions.

	Pretest/Invention Condition		Pretest/Worked Examples Condition	
	Post-Pretest	Post-Activity	Post-Pretest	Post-Activity
Cognitive Load	5.20 (1.39)	5.89 (1.57)	5.43 (1.96)	5.10 (1.97)
Knowledge Gaps	3.02 (0.88)	3.30 (0.79)	3.13 (0.95)	2.36 (0.86)
Interest and Enjoyment	3.24 (0.79)	3.38 (0.83)	3.34 (0.79)	3.51 (0.83)

Table 5
Means (standard deviations in parentheses) of posttest and post-activity questionnaire data as a function of pretest condition and activity type (Experiment 3).

	Pretest		No-Pretest	
	Invention	Worked Examples	Invention	Worked Examples
Posttest Score	2.57 (0.69)	2.98 (0.64)	2.79 (0.75)	2.68 (0.75)
Cognitive Load	5.81 (1.62)	5.22 (1.94)	5.68 (1.36)	5.41 (1.64)
Knowledge Gaps	3.30 (0.79)	2.40 (0.82)	3.55 (1.00)	2.76 (1.00)
Interest and Enjoyment	3.38 (0.81)	3.47 (0.82)	3.36 (0.69)	3.27 (0.86)

(1,142) = 2.44, $p = .121$, $\eta_p^2 = 0.02$, pretest condition, or interaction, $Fs < 1$.

For perceived knowledge gaps, a main effect of activity type was found, $F(1,142) = 31.52$, $p < .001$, $\eta_p^2 = 0.18$. Students who invented ($M = 3.43$, $SD = 0.90$) reported greater knowledge gaps than those who explored worked examples ($M = 2.59$, $SD = 0.93$). A main effect of pretest condition was also found, $F(1,142) = 4.21$, $p = .042$, $\eta_p^2 = 0.03$. Students who received a pretest ($M = 2.85$, $SD = 0.91$) experienced lower perceived knowledge gaps post-activity than those who did not ($M = 3.16$, $SD = 1.07$). There was no interaction, $F < 1$. Following the activity, both exploring worked examples and completing a pretest lowered perceived knowledge gaps, but these two factors did not have a significant combined impact.

For interest and enjoyment following the activity, no main effects or interaction were found, $Fs < 1$.

5.2.5. Conclusion

Confirming our hypothesis, completing a pretest led to higher posttest scores for students who explored using worked examples compared to those who invented. When no pretest was given, students learned at the same level in the worked examples and invention conditions. Although students in the pretest conditions received an additional questionnaire following the pretest, these results replicated those of Experiment 1 in which no questionnaire was given following the pretest. Thus, adding the questionnaire did not likely impact learning.

In the pretest/invention condition, cognitive load increased from post-pretest to post-activity. In contrast, cognitive load remained constant in the pretest/worked examples condition. Thus, the invention activity increased students' mental effort relative to the pretest. This increase may have been due to the greater number of elements in the activity, increased time, or completing two problems in a row. However, following the activity, cognitive load did not differ between the four conditions. We consider these findings, and those for the other questionnaires, more in the next section.

6. General discussion

Exploratory learning prior to direct instruction has been demonstrated to improve learning across several domains, age groups, and types of activities (see Alfieri et al., 2011; Kapur, 2016; Loibl et al., 2016). However, there is a need for more tightly controlled experimental conditions in this literature (Glogger-Frey et al., 2015; Hsu et al., 2015; Loibl et al., 2016; Schwartz, Chase, Opezzo, & Chin, 2011; Sweller, Kirschner, & Clark, 2007). In addition, not every study finds a benefit of exploring (e.g., Chase & Klahr, 2017; Fyfe, DeCaro, & Rittle-Johnson, 2014), indicating that more work is needed to understand the mechanisms supporting exploratory learning. Across three experiments, we examined whether reducing the cognitive demands of exploration enhances its learning benefits. Undergraduate students explored the topic of statistical variance, using materials adapted from previous exploratory learning studies.

In Experiment 1, students who explored worked examples (full guidance) prior to receiving instruction showed greater learning, lower cognitive load, and lower knowledge gaps than students in the invention condition (no guidance). Students in the completion problems condition (partial guidance) did not have significantly different scores than either other condition, demonstrating a middling effect. This benefit of worked examples to learning and cognitive load is consistent with previous research (Paas, 1992; Tuovinen & Sweller, 1999), and extends this work to an exploratory learning context.

In Experiment 2, students in a psychological statistics course completed invention or worked examples activities either as exploration (before instruction) or as practice (after instruction). There was an overall benefit of exploring over receiving instruction first, replicating and extending prior exploratory learning studies to undergraduate statistics using a tightly-controlled experimental design. However, we did not replicate the findings of Experiment 1—students who explored using worked examples showed identical learning outcomes to those who invented, despite reporting lower cognitive load and knowledge gaps.

Experiment 3 tested whether the use of a pretest explains these conflicting results. A pretest was given in Experiment 1 but not in Experiment 2. In Experiment 3, students were either given a pretest or not, then explored using either worked examples or an invention problem. Exploring with worked examples led to greater learning than inventing when students also completed a pretest, but not when the pretest was absent. These findings mirror the results of Experiments 1 and 2, respectively, and demonstrate that a pretest may serve as an invention activity. Although suggested by others (e.g., Glogger-Frey et al., 2017, 2015; Kapur, 2016), the impact of a pretest in exploratory learning research has not been systematically tested.

6.1. Using a pretest in learning experiments

The pretest may have activated similar learning mechanisms thought to underlie exploratory learning, even if answered incorrectly. These mechanisms mirror those suggested by the literature on testing effects more generally. The process of retrieval allows students to

activate and reconstruct relevant prior knowledge, and can have a potentiating effect whereby new information is better encoded in memory (Carpenter, 2011; Kapur, 2012; Karpicke, 2012; Richland et al., 2009; Rowland, 2014). Pretesting can also draw students' attention to the most important features (Karpicke, 2012; Richland et al., 2009). Finally, pretesting can improve metacognition, by eliciting awareness of knowledge gaps (Rohrer & Pashler, 2010).

The finding that the pretest impacted learning has widespread implications for research comparing direct instruction to constructivist-inspired teaching methods such as exploratory learning. For example, Klahr and Nigam (2004) compared pure discovery learning to a direct instruction condition and found greater learning in the direct instruction condition. However, Klahr and Nigam also used an extensive pretest which targeted the to-be-learned material. This pretest may have activated the mechanisms thought to underlie exploratory learning. Thus, their direct instruction condition is better characterized as an exploratory learning condition. Using this framing, Klahr and Nigam actually demonstrated that exploratory learning (not direct instruction) led to higher learning outcomes than pure discovery learning.

A similar conclusion could be applied to a study by Chase and Klahr (2017), who ostensibly compared exploratory learning to a tell-then-practice condition and found no differences in learning outcomes. However, in both conditions, a 30-min pretest was given in a prior session. An additional pretest was given at the beginning of the intervention session, immediately preceding the exploration and direct instruction activities. If the pretest served as an exploration activity, these null results are less surprising. Chase and Klahr simply compared two exploratory learning conditions, with one versus multiple exploration activities given prior to instruction.

Use of a pretest might also explain some of the mixed results in previous studies comparing guided and unguided exploration. Glogger-Frey et al. (2015) used a pretest in the first experiment, and not in the second. Although they found a benefit of exploring with worked examples over invention in both experiments, this benefit was restricted to far transfer in the second experiment. Likourezos and Kalyuga (2016) used a pretest that only included items assessing prerequisite knowledge, not knowledge of the targeted concepts. They found comparable learning between unguided and guided exploration conditions, although there were more creative and advanced solutions provided in the guided exploratory learning condition. Glogger-Frey et al. (2017) did not use a pretest, and also gave additional practice on the activity in both conditions, and found the reverse effect—that exploring with invention led to better far transfer than exploring with worked examples. These results roughly align with the idea that including a pretest with guided exploration improves learning relative to unguided exploration, and not including a pretest leads to little or no benefit of guidance.

6.2. Mechanisms of exploratory learning

Taken together, our results provide important information about the mechanisms by which exploration activities benefit learning. By comparing the impact of guidance and a pretest on learning outcomes and survey measures, we conclude that successful exploration activities serve at least two key functions: increasing perceived knowledge gaps, and enabling students to discern relevant problem features. Activating prior knowledge and increasing situational interest and enjoyment may play a role as well, but our studies either did not systematically test these factors or show differences between conditions.

6.2.1. Cognitive load, knowledge gaps, and problem features

Across the experiments, increasing guidance during exploration by using worked examples lowered reported cognitive load, although this finding was not statistically significant in Experiment 3. However, worked examples did not consistently improve learning: When there was no pretest before the worked examples, learning did not improve relative to the invention condition. Thus, reducing cognitive load is not sufficient to improve learning.

However, worked examples did improve learning when combined with a pretest, demonstrating that the pretest served an important additional function. Our findings indicate that the pretest increased knowledge gaps. In Experiment 3, we assessed knowledge gaps immediately following the pretest. In the worked examples condition, knowledge gaps decreased from pretest to activity. In the invention condition, knowledge gaps remained at the same, higher, level. Thus, knowledge gaps were instantiated by the pretest, but resolved by the worked examples. Across all studies, knowledge gaps were consistently higher after the activity in the invention conditions than in the worked examples conditions. Invention conditions also led to lower learning.

Together, these cognitive load and knowledge gap findings suggest that an initial experience that increases knowledge gaps, combined with an exploration activity that provides guidance, improves exploratory learning. Without knowledge gaps, reducing cognitive load has little benefit, as shown in the no-pretest/worked examples conditions. And without supporting cognitive load, knowledge gaps have little advantage, as demonstrated in the invention conditions.

We further propose that worked examples supported important cognitive learning mechanisms. Inventing a problem solution can be difficult for students (Kapur, 2015). In contrast, worked examples constrain the solution possibilities, possibly helping students discern the structural problem features from the superficial features (Glogger-Frey et al., 2015). Following, direct instruction may help students build on such knowledge and allow for a better connection between the procedures they viewed during exploration and the conceptually rich instruction.

Thus, combining a pretest with guidance prior to more direct instruction may benefit learning via both metacognitive and cognitive mechanisms. This latter mechanism is largely overlooked by Kalyuga and Singh (2016), who primarily discussed the metacognitive mechanisms of exploratory learning, and argued that it may be irrelevant to consider cognitive load during exploration. Our findings suggest that reducing cognitive load during exploration directly increases domain-relevant knowledge and/or conceptual understanding and transfer. Therefore, considering cognitive load may help instructors design better exploration activities.

6.2.2. Other mechanisms

Exploratory learning is also thought to activate prior knowledge, so that new information from instruction can be better integrated with already-existing schemas in long-term memory (Kapur, 2015; Schwartz et al., 2007). This mechanism seems inherent to all exploration activities, as students must draw upon their knowledge to grapple with a new concept or problem prior to instruction. However, we found that some exploratory learning conditions were better than others. Thus, activating prior knowledge is unlikely the sole mechanism driving the benefits of exploratory learning.

Situational interest and enjoyment is also unlikely a driver of our results, as this measure consistently showed no differences between conditions. Importantly, this finding also demonstrates that neither the more proscribed worked examples nor the more taxing invention activity dampened students' interest. The means were consistently slightly above the scale midpoint, suggesting that students found the activities somewhat interesting. Previous studies have found mixed results on interest and enjoyment measures (Glogger-Frey et al., 2015; Weaver et al., 2018). However, these null results are inconsistent with work showing that perceiving more knowledge gaps enhances interest and curiosity (Glogger-Frey et al., 2015; Rotgans & Schmidt, 2014). In our study, interest was not higher in the invention conditions, when knowledge gaps were also higher. One possible reason for this inconsistency is that students in our study were asked to rate how interesting they perceived the learning activity, whereas previous studies assessed how interested in or curious about the topic they were (Glogger-Frey et al., 2015). Perceived knowledge gaps may not enhance interest in the

learning activity itself, but rather lead to interest in seeking elsewhere for information to fill these gaps.

6.3. Methodological considerations and limitations

These conclusions must be considered in the context of our methodology. More research is needed to determine if these results generalize to other types of students (e.g., children) or learning domains. In addition, our findings may have been different if we had used (a) contrasting cases in our exploration activity, (b) a different type of pretest, or (c) different posttest items. We also cannot definitively speak to whether our use of worked examples functioned as exploration or instruction.

6.3.1. Contrasting cases

We argue that worked examples benefit exploratory learning by helping to make the problem structure more apparent. This support may have been especially helpful because we used rich datasets, which present data tables with a lot of information that students must work with—although we only used six data points for each of the three tea growers in our activity. Many other studies instead use contrasting cases (e.g., Glogger-Frey et al., 2017, 2015; Loibl & Rummel, 2014b; Schwartz et al., 2011; Schwartz & Martin, 2004). With contrasting cases, students compare and contrast across cases that differ by a single problem feature, highlighting the deep structure of the problem. In our activity, one dataset had a feature that contrasted the other two (i.e., it was missing a value, to highlight the importance of sample size). However, the other two datasets did not differ in any significant way. By highlighting important problem features, both worked examples and contrasting cases may serve comparable goals. However, Glogger-Frey et al. (2015) used contrasting cases in their exploration activities, and found that worked examples still enhanced learning and reduced cognitive load compared to inventing.

6.3.2. Type of pretest

We found that combining a pretest with worked examples enhanced knowledge gaps and learning. However, the type of pretest items may impact whether knowledge gaps increase, or whether they benefit exploratory learning. If our pretest had only assessed central tendency, students may have had directed less motivation or attention to the topic of consistency. Consistent with this idea, Likourezos and Kalyuga (2016) used a pretest assessing prerequisite knowledge, not knowledge of the target concepts. They found comparable learning between unguided and guided exploratory learning conditions. Findings might also differ for open-ended versus closed-ended items such as multiple choice questions (cf. Richland et al., 2009).

6.3.3. Posttest subscales

Most exploratory learning studies find benefits to conceptual knowledge and/or transfer, rather than procedural fluency. We did not find condition differences as a function of subscale. One possibility is that our measure conflates these three subscales. For example, for students to receive full credit on our procedural fluency item, they must not only compute standard deviation but also understand what the resulting value means conceptually. Future research is needed using items that more fully differentiate these knowledge types.

6.3.4. Is use of worked examples exploration or instruction?

We have argued that worked examples function as exploration when used prior to more detailed procedural and conceptual instruction. Although students are guided through the procedures for solving a novel problem, they have not been told the conceptual basis of the problem. However, this framing is debatable, and worked examples may simply be an additional form of direct instruction. If this were the case, then one might argue that the pretest served as the invention activity in the pretest/worked examples conditions. In this case, our findings would demonstrate that providing a very brief, open-ended pretest on the target concept followed by more extensive direct instruction led to better performance than the same pretest followed by a more extensive invention activity. More research is needed to determine whether the shortened pretest would have served the same role as invention. In addition, worked examples can include more conceptual information than ours did. Adding this information could lead to even less exploration, and findings more akin to a tell-then-practice condition.

6.4. Conclusion

Despite the benefits of constructivist-inspired teaching methods, not all of these methods are necessarily better than direct instruction. The current findings provide greater insight into how a particular type of constructivist method—exploratory learning—may best benefit student understanding. Exploration activities should be designed to help students perceive gaps in their knowledge, but also support cognitive load. By doing so, students may be best able to draw out the deep structure of the problem, supporting conceptual understanding.

Acknowledgements

The authors thank Michael Wiedmann for providing his experiment materials. Portions of this work were presented at the 2018 meeting of the Cognitive Science Society. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.learninstruc.2019.05.005.

References

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology, 103*(1), 1–18. https://doi.org/10.1037/a0021017.supp.

Berlyne, D. E. (1954). A theory of human curiosity. *British Journal of Psychology, 45*(3), 180–191.

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.

Borek, A., McLaren, B. M., Karabinos, M., & Yaron, D. (2009). How much assistance is helpful to students in discovery learning? In U. Cress, V. Dimitrova, & M. Specht (Eds.). *Proceedings of the fourth european conference on technology enhanced learning, learning in the synergy of multiple disciplines (EC-TEL 2009), LNCS 5794, september/ october 2009, nice, France* (pp. 391–404). Springer-Verlag Berlin Heidelberg.

Carpenter, S. K. (2011). Semantic information activated during retrieval contributed to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 1547–1552. https://doi.org/10.1037/a0024140.

Chase, C. C., & Klahr, D. J. (2017). Invention versus direct instruction: For some content, it's a tie. *Journal of Science Education and Technology, 26*, 582–596. https://doi.org/10.1007/s10956-017-9700-6.

Chen, O., Kalyuga, S., & Sweller, J. (2015). The worked example effect, the generation effect, and element interactivity. *Journal of Educational Psychology, 107*(3), 689–704. https://doi.org/10.1037/edu0000018.

Dean, D., Jr., & Kuhn, D. (2007). Direct instruction vs. discovery: The long view. *Science Education, 91*, 384–397.

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. https://doi.org/10.1016/j.learninstruc.2011.08.003.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160. https://doi:10.3758/BRM.41.4.1149.

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

Fyfe, E. R., DeCaro, M. S., & Rittle-Johnson, B. (2014). An alternative time for telling: When conceptual instruction prior to problem solving improves mathematical

knowledge. *British Journal of Educational Psychology, 84*, 502–519. https://doi.org/10.1111/bjep.12035.
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.
Glogger-Frey, I., Gaus, K., & Renkl, A. (2017). Learning from direct instruction: Best prepared by several self-regulated or guided invention activities? *Learning and Instruction, 51*, 26–35.
Hsu, C.-Y., Kalyuga, S., & Sweller, J. (2015). When should guidance be presented during physics instruction? *Archives of Scientific Psychology, 3*, 37–53.
Jarosz, A. F., Goldenberg, O., & Wiley, J. (2016). Learning by invention: Small group discussion activities that support learning in statistics. *Discourse Processes, 54*(4), 285–302. https://doi.org/10.1080/0163853X.2015.1129593.
Kalyuga, S., & Singh, A. (2016). Rethinking the boundaries of cognitive load theory in complex learning. *Educational Psychology Review, 28*, 831–852. https://doi.org/10.1007/s10648-015-9352-0.
Kapur, M. (2012). Productive failure in learning the concept of variance. *Instructional Science, 40*(4), 651–672.
Kapur, M. (2014). Productive failure in learning math. *Cognitive Science, 38*(5), 1008–1022.
Kapur, M. (2015). Learning from productive failure. *Learning: Research and Practice, 1*, 51–65. https://doi.org/10.1080/23735082.2015.1002195.
Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educational Psychologist, 51*(2), 289–299. https://doi.org/10.1080/00461520.2016.1155457.
Karpicke, J. D. (2012). Retrieval-based learning: Active retrieval promotes meaningful learning. *Current Directions in Psychological Science, 21*, 157–163. https://doi.org/10.1177/0963721412443552.
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. https://doi.org/10.1207/s15326985ep4102_1.
Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science, 15*, 661–667.
Likourezos, V., & Kalyuga, S. (2016). Instruction-first and problem-solving-first approaches: Alternative pathways to learning complex tasks. *Instructional Science*. https://doi.org/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*. https://doi.org/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.
Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist, 59*(1), 14–19. https://doi.org/10.1037/0003-066X.59.1.14.
Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43–52.
van Merriënboer, J., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 12*(2), 147–177. https://doi.org/10.1007/s10648-005-3951-0.
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. https://doi.org/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.
Richland, L. E., Kornell, N., & Kao, L. E. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied, 15*(3), 243–257. https://doi.org/10.1037/a0016496.
Rohrer, D., & Pashler, H. (2010). Recent research on human learning challenges conventional instructional strategies. *Educational Researcher, 39*, 406–412. https://doi.org/10.3102/0013189X10374770.
Roll, I., Aleven, V., & Koedinger, K. R. (2009). Helping students know 'further'—-increasing the flexibility of students' knowledge using symbolic invention tasks. In N. A. Taatgen, & H. van Rijn (Eds.). *Proceedings of the 31st annual conference of the cognitive science society* (pp. 1169–1174). Austin: Cognitive Science Society.
Rotgans, J. I., & Schmidt, H. G. (2014). Situational interest and learning: Thirst for knowledge. *Learning and Instruction, 32*, 37–50. https://doi.org/10.1016/j.learninstruc.2014.01.002.
Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin, 140*(6), 1432–1463.
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., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction, 16*(4), 475–522.
Schwartz, D. L., Chase, C. C., Opezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology, 103*, 759–775. https://doi.org/10.1037/a0025140.
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.
Schwartz, D. L., Sears, D., & Chang, J. (2007). Reconsidering prior knowledge. In M. Lovett, & P. Shah (Eds.). *Thinking with data* (pp. 319–344). Mahwah, NJ: Lawrence Erlbaum.
Silvia, P. J. (2008). Interest—the curious emotion. *Current Directions in Psychological Science, 17*(1), 57–60.
Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory, 4*, 592–604. https://doi.org/10.1037/0278-7393.4.6.592.
Sweller, J. (2004). Instructional design consequences of an analogy between evolution by natural selection and human cognitive architecture. *Instructional Science, 32*(1-2), 9–31. https://doi.org/10.1023/B:TRUC.0000021808.72598.4d.
Sweller, J., Kirschner, P., & Clark, R. E. (2007). Why minimally guided teaching techniques do not work: A reply to commentaries. *Educational Psychologist, 42*, 115–121. https://doi.org/10.1080/00461520701263426.
Sweller, J., van Merriënboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296.
Tuovinen, J. E., & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology, 91*, 334–341.
Weaver, J. P., Chastain, R. J., DeCaro, D. A., & DeCaro, M. S. (2018). Reverse the routine: Problem solving before instruction improves conceptual knowledge in undergraduate physics. *Contemporary Educational Psychology, 52*, 36–47. https://doi.org/10.1016/j.cedpsych.2017.12.003.
Wiedmann, M., Leach, R. C., Rummel, N., & Wiley, J. (2012). Does group composition affect learning by invention? *Instructional Science, 40*(4), 711–730.
Wiedmann, M., Leach, R. C., Rummel, N., & Wiley, J. (2015). Mathematical skills and learning by invention in small groups. In Y. H. Cho, S. I. Caleon, & M. Kapur (Eds.). *Authentic problem solving and learning in the 21st century: Perspectives from Singapore and beyond* (pp. 249–265). Singapore: Springer Science + Business Media.
Wise, A. F., & O'Neill, K. (2009). Beyond more versus less: A reframing of the debate on instructional guidance. In S. Tobias, & T. M. Duffy (Eds.). *Constructivist instruction: Success or failure?* (pp. 82–105). New York, NY: Routledge/Taylor & Francis Group.