Educational Psychology Review (2026) 38:12
https://doi.org/10.1007/s10648-025-10103-6

INTERVENTION STUDY

Conditions for Effective Learning Without Upfront Instruction: How Practice with Feedback Supports Memory, Generalization, Motivation, and Metacognition

Michael W. Asher[1] · **Paulo F. Carvalho**[1]

Received: 26 July 2025 / Accepted: 9 December 2025

Abstract

What conditions are necessary for students to learn from practice and feedback, without the need for upfront lecture? Across two experiments ($N = 597$), we examined how practice with feedback can support memory, generalization, metacognition, and motivation. Participants were randomly assigned to one of three instructional formats: a traditional lecture, practice with correct-answer feedback, or practice with explanatory feedback (predefined or adaptive and AI generated). In both studies, the lecture condition introduced linear regression through definitions and a worked example, while the practice conditions used matched problem sets with feedback that either (a) provided only correct answers or (b) explained why answers were correct. Study 1 used multiple-choice questions; Study 2 used open-ended questions with personalized explanatory feedback generated in real time by GPT-4o. For memory, both types of feedback outperformed lecture, suggesting that attempting a response and receiving feedback—even without explanations—enhances encoding. For generalization, however, feedback needed to include explanations, and learners needed sufficient prior knowledge to benefit. Study 2 also showed that practice—regardless of feedback type—improved metacognitive calibration compared to lecture, helping learners more accurately assess their understanding. While lecture produced greater situational interest for less-confident learners in Study 1, this pattern reversed in Study 2, where personalized, AI-generated feedback elicited higher interest for this group. Together, these findings clarify when and for whom practice with feedback can replace lecture-based instruction, and they highlight the potential of generative AI to scale personalized, explanatory feedback.

Keywords Feedback · Practice testing · Active learning · Generative AI

Author Note All data and code for this study are openly available at https://osf.io/kruzw. Hypotheses, methods, and the analysis plan for Studies 1 and 2 were pre-registered at https://osf.io/j4rz6 and https://osf.io/28jtr.

Extended author information available on the last page of the article

Published online: 26 January 2026

In education, it is widely assumed that students must first receive explicit instruction—typically through lectures or readings—before they are ready to engage in active learning. This belief shapes how teachers are trained and how classrooms are designed. For example, math and science educators are often taught to follow a gradual release model—"I do, we do, you do"—that delays guided and independent practice until after teacher-led explanation (Fisher & Frey, 2021). Similarly, college courses tend to frontload lectures and readings, with hands-on activities and discussion occurring later (Stains et al., 2018). In medical education, the adage "see one, do one, teach one" likewise implies that observation should precede active participation (Rodriguez-Paz et al., 2009).

This belief—that passive instruction must precede active engagement—also underlies much of the learning sciences literature. Many studies on retrieval practice compare passive instruction alone (e.g., reading or lecture) to passive instruction followed by an active task (e.g., practice testing). Implicit in these designs is the belief that students must first be told or shown information before they can benefit from engaging with it. For example, in a classic study by Roediger and Karpicke (2006), participants read short passages and were then assigned either to re-read or to recall and summarize them from memory. In related work on successive relearning, Rawson et al. (2013) had students study psychology vocabulary—terms they had previously encountered through instruction—using different schedules of spaced retrieval. Across these studies, memory consistently improved when active testing followed prior instruction.

But is upfront, passive instruction truly necessary for active strategies like practice testing to be effective? Could students benefit from engaging in active learning before receiving traditional instruction—or even in place of it?

Emerging research suggests that the timing and format of instruction may be more flexible than traditionally assumed. Studies on problem-solving before instruction show that engaging with challenging problems—even before receiving explicit teaching—can enhance subsequent learning (DeCaro & Rittle-Johnson, 2012; Loibl & Rummel, (2014). In pre-questioning studies, students who attempt questions before a lecture often remember more than peers who simply listen (Carpenter et al., 2018; Little & Bjork, 2016). In productive failure designs, students typically struggle with ill-structured problems across multiple class periods before teachers consolidate learning through discussion and explanation (e.g., Kapur & Bielaczyc, 2012). Although initial problem-solving is often unsuccessful, the combination of active exploration followed by direct instruction produces stronger conceptual understanding than instruction alone.

These findings sit at the intersection of a longstanding debate between constructivist and direct instruction approaches. Constructivists argue that learners develop understanding by actively discovering principles themselves (Bruner, 1961). Proponents of direct instruction counter that novices lack the existing knowledge structures needed to make sense of new information during unguided exploration, making explicit teaching more efficient and effective (Kirschner et al., 2006; Sweller, 2004). Problem-solving before instruction occupies a middle ground: students engage in discovery-like exploration but then immediately receive direct instruction in the form of feedback to consolidate correct understanding.

Asher, Sana, et al. (2025) tested whether practice with targeted feedback could have the potential to replace traditional instruction altogether. If pre-questions help students learn more from lectures, they reasoned, practice questions should also help students learn directly from feedback. In three studies, students who learned statistics via practice and feedback—without any lecture—performed as well as their peers in the lecture condition and learned two to three times faster. Importantly, in Studies 2 and 3, these students didn't just memorize answers—they successfully applied what they learned to novel problems involving linear regression (Study 2) and measures of central tendency (Study 3).

This work suggests that practice with feedback represents a potentially efficient variant of the middle ground between constructivist and direct instruction approaches. Like productive failure and pre-questioning, it begins with student problem-solving. But unlike productive failure, it does not defer instruction until after extended struggle, and unlike pre-questioning, it avoids the need for extended lecture; instead, it delivers immediate and concise feedback after each attempt, blending the active learning of discovery with the explicit guidance of instruction. Rather than separating exploration and explanation, practice-centered instruction offers these experiences iteratively and tightly coupled.

Still, these findings raise a critical question: under what conditions can practice with feedback support both memory and conceptual understanding? Asher et al. demonstrated that it can work but left open questions about when and why. To identify mechanisms and boundary conditions that determine success, we propose a theory of practice-centered learning that integrates theoretical insights from research on problem-solving before instruction, including work on productive failure, pre-questioning, and replacing lecture with practice and feedback. Based on these insights, we suggest that practice with simple correct-answer feedback will enhance memory and metacognitive awareness relative to lecture, but that successful generalization from practice alone requires more: feedback must explain underlying principles, and learners must have sufficient background knowledge to engage productively with these explanations.

A Theory of Practice-Centered Learning

This theory begins with the premise that effective learning requires two elements: (1) learners must receive clear instruction, and (2) learners must be prepared to encode and retain this instruction effectively. When only one element is present, learning suffers. This principle applies to all forms of instruction, including lecture and feedback.

We argue that lecture and feedback can be equivalent in the instructional content they provide: well-designed feedback can deliver the same key information as a lecture, whether this includes facts to be memorized or the rules and principles required to generalize. However, lecture and feedback should differ in how well they prepare learners to encode and retain instruction. When feedback is paired with upfront practice, this active task should engage attention and sharpen metacognitive monitoring, better preparing students to encode and retain the instructional content that feedback provides (Asher, Sana, et al., 2025; Carvalho et al., 2024; Pan & Carpenter, 2023).

This preparation mechanism—observed across productive failure, pre-questioning, and practice-with-feedback designs—should support both memory and generalization. However, whereas memory may benefit from practice and feedback regardless of feedback quality, successful generalization should depend on two key ingredients: feedback must explain underlying principles, and learners must have sufficient prior knowledge to engage productively with these explanations. When these ingredients are absent, practice may enhance retention without supporting the deeper understanding needed to apply knowledge flexibly.

At the same time, because practice with feedback increases challenge, it may trigger curiosity and provide opportunities for students to experience success and feel efficacious. However, this challenge may also feel frustrating or discouraging, particularly for learners with lower confidence or less prior knowledge. Table 1 summarizes the known or likely affordances of practice, lecture, and feedback for metacognition, memory, generalization, and motivation.

Metacognitive Monitoring

Practice with feedback should effectively support metacognitive monitoring—in which people reflect on their own cognitive and memory processes (Koriat, 2007). Lectures typically offer little opportunity for students to monitor their knowledge (Chi & Wylie, 2014) and may foster overconfidence when students mistake fluent comprehension for true understanding (Bjork & Bjork, 2011). In contrast, practice with feedback provides ongoing self-assessment opportunities through immediate signals—both internal (ease of retrieval) and external (answer correctness)—that help students identify what they know and don't know (Koedinger et al., 2012).

Metacognitive monitoring encompasses multiple dimensions, including subjective judgments of learning (assessments of how well one has learned; Nelson & Narens, 1990) and calibration (the accuracy of predicted versus actual performance; Schraw, 2009). Practice with feedback may influence these dimensions differently. While it should improve calibration by providing concrete feedback about performance, it may paradoxically lower judgments of learning as students who struggle through practice problems become more aware of knowledge gaps—even when practice leads to better actual preparation than lecture. Thus, while practice with feedback may not affect how much students feel they learned (Asher, Sana, et al., 2025), it might decrease overconfidence relative to passive instruction and make learners more accurate in their self-assessments.

Memory

Practice with feedback should consistently outperform lecture for promoting memory. When students attempt practice problems and receive feedback, several processes enhance encoding and retention. First, the act of attempting problems activates specific learning goals and directs attention toward understanding correct responses (Carvalho et al., 2024; Pan & Carpenter, 2023). Second, retrieving information from memory strengthens retention more than passive study (Roediger & Karpicke, 2006). Finally, receiving feedback about correctness helps students identify what they know

Table 1 Affordances of Practice, Lecture, and Feedback for Different Outcomes

Outcome	Practice	Lecture	Feedback
Metacognition	(+) Can inform self-evaluation, helping to identify gaps in knowledge.[1]	(−) Without additional scaffolds such as reflection prompts, it offers limited support for metacognition.[2]	(+) Directly informs the learner about known information and knowledge gaps.[1]
Memory	(+) Enhances long-term memory via retrieval practice.[3]	(−) Learners can commit lectured information to memory if they watch attentively. Due to the passive nature of lecture, this can be a challenge.[2] (+) Providing explanations (e.g., the origins of a word or rationale behind a procedure) can promote deeper semantic processing and thereby improve recall.[4]	(+) Learners can commit feedback to memory if they attend to it. Because of the active nature of practice and feedback and its benefits for metacognitive calibration, attention should be likely.[5] Explanatory feedback can communicate the same semantic cues as lecture.
Generalization	(+) Varied practice can support induction of underlying rules and principles needed for generalization, if students can handle the cognitive load of this demanding task.[6,7]	(+) Explanations of rules and principles can shortcut the inductive reasoning process, allowing for generalization with less cognitive load.[8]	(+) Explanatory feedback should have the same benefits as lecture if it explains rules and principles.[9]
Motivation	(+) Success can increase confidence and perceived progress can build self-efficacy.[10] Challenge can promote situational interest in the lesson.[11] Relevant practice problems can promote interest in the content.[12] (−) Repeated failure can decrease confidence and undermine intrinsic motivation.[13]	(+) Dynamic, engaging instruction can trigger situational interest. Discussion of relevance can promote interest in the content.[11,14,15]	(+) Supportive feedback may be able to buffer students from motivational harms of failure during practice. Harsh feedback can further undermine intrinsic motivation.[13,16]

[1]Koedinger et al., 2012, [2]Chi & Wylie, 2014, [3]Roediger & Karpicke, 2006, [4]Craik & Tulving, 1975, [5]Carvalho et al., 2024, [6]Raviv et al., 2022, [7]Kirschner et al., 2006, [8]Sweller, 2004, [9]Asher, Sana, et al., 2025, [10]Bandura, 1997, [11]Renninger et al., 2019, [12]Bernacki & Walkington, 2018, [13]Ryan & Deci, 2000, [14]Durik & Harackiewicz, 2007, [15]Asher & Harackiewicz, 2024, [16]Anderson & Rodin, 1989

and don't know, supporting both initial encoding and self-directed review (Koedinger et al., 2012). In contrast, lectures may fail to activate these processes unless students are already highly motivated and strategic.

The advantages of practice and feedback for memory should hold even when feedback consists solely of correct answers: this minimal level of support allows students to evaluate their accuracy and commit correct information to memory (e.g., Butler et al., 2007). Although explanations in lectures or elaborated feedback can promote semantic processing and improve memory (Craik & Tulving, 1975), we anticipate that even simple correct-answer feedback should often yield greater memory gains than lectures through its support for focused attention and active engagement.

Generalization

Generalization—the ability to apply learned principles to novel problems—is a higher-order learning outcome that requires understanding the underlying structure that governs a domain (Bloom et al., 1956; Koedinger et al., 2012; Krathwohl, 2002). Here, we do not expect that all forms of practice and feedback will outperform lectures. Instead, successful generalization should depend on two critical ingredients: explanatory feedback and prior knowledge.

Explanatory Feedback Lectures are often thought to hold a generalization advantage because they directly convey abstract principles, reducing the cognitive load associated with discovering them (Sweller, 2004). However, practice-centered instruction should be able to achieve similar benefits when materials make underlying principles explicit—whether through initial practice problems targeting key definitions and formulas, or through feedback that explains rationales behind correct answers. These approaches should preserve the benefits of active engagement while scaffolding students' understanding of underlying principles.

Supporting this prediction, a meta-analysis of 77 experimental studies on computer-based feedback found that elaborated feedback was more than twice as effective as correct-answer feedback for promoting higher-order learning outcomes (g=.46 vs. g=.22, Mertens et al., 2022). In addition, Asher and colleagues (2025) used practice and feedback that explained underlying rules to promote generalization at or above the level of lecture-based instruction.

Prior Knowledge Even well-crafted explanatory feedback may fail to support generalization if learners lack the background knowledge needed to make sense of it. For learners with strong foundational knowledge, a brief correction may be enough to infer a relevant principle. For novices, however, explanation-heavy feedback may overwhelm working memory and provide little instructional value (Kirschner et al., 2006). Practice and feedback may help these learners recognize gaps in their knowledge, but without adequate scaffolding or foundational knowledge, they may be unable to resolve the confusion.

Motivation

In many respects, the motivational affordances of lecture- and practice-centered instructional approaches may be quite similar. For instance, a lecture might spark situational interest by focusing on a fun or personally relevant topic—but so can a well-designed problem set (Bernacki & Walkington, 2018; Walkington, 2013). Similarly, instructors can foster deeper engagement by discussing the relevance of course content during a lecture (Asher & Harackiewicz, 2024), but they should also be able to incorporate the same information into elaborated feedback or practice materials. Still, two important motivational differences are likely to emerge. First, the challenge of solving practice problems can serve as a powerful trigger of situational interest—a temporary state of heightened attention and affective engagement (Renninger et al., 2019). Second, the frequent correctness feedback sends strong signals about competence and potential—beliefs that are central to all major theories of motivation (e.g., Eccles & Wigfield, 2020; Ryan & Deci, 2000). Success can improve a learner's beliefs about their ability to succeed and overcome challenges (Bandura, 1997) but repeated failure can decrease confidence and undermine intrinsic motivation (Ryan & Deci, 2000). Moreover, initial confidence may serve as a moderator for motivational outcomes: students with higher confidence might appreciate the challenge of practice without upfront lecture, while less-confident students might feel overwhelmed (Asher, Sana, et al., 2025; Zepeda et al., 2020).

The Present Research

In the present research, we examined predictions derived from our theory. First, we tested two theory-derived boundary conditions for when practice with feedback will produce comparable or superior learning outcomes to lecture: (1) whether students are asked to memorize information or form new generalizations, and (2) whether feedback provides simple correct answers or explanations of underlying principles. We also examined moderators (prior knowledge/confidence) and additional consequences (judgments of learning, metacognitive calibration, interest, and efficiency) of learning through practice with feedback.

We tested these predictions in two studies. In Study 1, we compared a lecture to multiple-choice practice questions with different types of feedback: either simple correct answers or explanations of underlying rules and principles. In Study 2 we ensured that our findings could generalize to open-ended practice and test questions, using Generative AI (GPT-4o) to provide personalized explanations.

Study 1

In Study 1, we conducted a three-cell online experiment that compared a statistics lecture (the control condition) with a matched set of practice problems and two forms of feedback: (1) *correct-response feedback*, consisting of only "correct" or "incorrect" plus the correct answer, and (2) *explanatory feedback*, which also includes a rationale for the correct answer.

For this study, we predicted that although practice with either type of feedback would be more effective than the lecture at promoting memory for factual information (H1), explanations would be needed for participants to generalize (H2). We also predicted that both forms of practice would take less instruction time (H3). Finally, we were concerned that participants, particularly those with less confidence in their math abilities, might be discouraged by the difficulty of solving practice problems without upfront guidance, struggling to learn and demonstrating decreased interest in the learning session or statistics content (H4).

To avoid pre-questioning effects—which can transform passive lecture into active learning (Pan & Carpenter, 2023)—we did not administer a statistics pretest in this initial study. As a result, we could not directly test predictions about the moderating role of prior knowledge for generalization performance. However, because confidence and prior knowledge are typically correlated (Wang, 2012 found r=.35–.44 in a 3000-student longitudinal study; Li et al., 2023 found r=.41 in a cross-sectional, representative sample of over 100,000 high schoolers), we explored whether the generalization benefits of explanatory feedback would be contingent on students' initial math confidence. To capture the impact of practice with feedback on metacognitive monitoring, we assessed students' subjective judgments of learning. Because we expected that practice with feedback would increase overall learning but also cue students to gaps in their knowledge, we made no directional predictions about this outcome.

This study was approved by the Institutional Review Board at Carnegie Mellon University. Hypotheses, along with data collection and analysis procedures for this study were pre-registered at https://osf.io/j4rz6.

Method

Participants

Data were collected online, with participants recruited via *Prolific*, a crowdsourcing platform. Participants were screened to be ≥18 years old and living in the United States. Following the preregistration, participants were recruited until 300 consented and completed the study. After excluding three participants (one for skipping the final test and two for submitting identical answers) the final sample was 297. Overall, 52% self-reported their ethnicity as White, 38% as Black, 6% as multiracial, 3% as Asian, and 2% as belonging to another group. 46% identified as women and 54% as men. The average age was 35.5 years. In terms of math background, 34% reported completing only high school-level math, 13% took an advanced placement math course, and 64% completed a college-level math course. The study lasted approximately 20 minutes, and participants were paid $4.00 for their participation.

Procedure

Study 1 was administered via *Qualtrics*. The study used a between-subjects design in which participants were randomly assigned to one of three instructional conditions: (1) a lecture, (2) practice with correct-response feedback, or (3) practice with explanatory feedback. Participants completed four sections in sequence: a baseline questionnaire, a statistics lesson, an outcome questionnaire, and a posttest.

At the beginning of the study, participants were informed they would learn about linear regression and take a test, then completed a baseline questionnaire rating their confidence in mathematics using a three-item scale. Next, participants proceeded to the statistics lesson, adapted from Asher, Sana, et al. (2025) and Asher and Harackiewicz (2024), covering six learning objectives related to linear regression: identifying the intercept of a line, calculating its slope, interpreting the slope, interpreting the intercept, forming a linear equation, and using this equation to make a prediction. In the lecture condition, participants watched an eight-minute video in which an instructor defined key terms, walked through a seven-step worked example (making two predictions and demonstrating each of the other five learning objectives once), and summarized the lesson. Approximately 45% of the lecture focused on definitions, 50% on the worked example, and 5% on summary.

In the two practice conditions, participants completed seven multiple-choice problems that exactly matched the seven steps in the video's worked example, with one practice problem per learning objective (except for predictions, which had two problems, mirroring the video). In the correct-response feedback condition, participants saw whether their answer was correct and were shown the correct answer. For example, after interpreting the slope of a line, a participant might have been told:

> **Incorrect**
> **Your answer:** 1.18 is the average murder rate.
> **The correct answer:** For every 1 degree increase in average temperature, the murder rate increases by 1.18 units.

Critically, these participants were not given a rationale for the correct answer. In contrast, in the explanatory feedback condition, participants received a detailed explanation of why the correct answer was correct, including relevant definitions and visual examples (see Fig. 1 for an example). This feedback was designed to mirror the conceptual content of the lecture in a more interactive format.

Following the learning session, participants completed an outcome questionnaire assessing their interest in statistics, their experience during the learning session, and

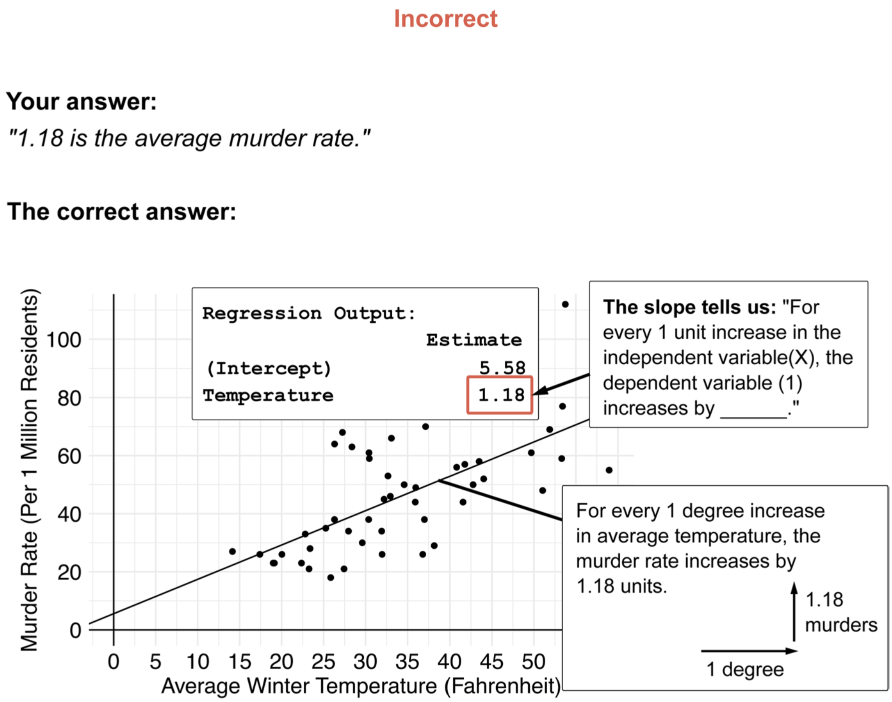

Fig. 1 Sample Explanatory Feedback, Study 1

their confidence and expectations for the posttest. Finally, all participants completed a 12-item regression test, which included three multiple-choice memory questions that were covered during instruction (in the worked example of the lecture or in a practice question) and nine short-answer generalization questions requiring application of learned concepts to novel problems. Short-answer questions were scored using a rubric shared in the Supplemental Materials.

Measures

All questionnaire measures used either five-point Likert-type scales (ranging from "Not at all" to "Extremely") or six-point agreement scales ("Strongly disagree" to "Strongly agree"). Baseline math confidence was assessed with three items (e.g., "How good are you at math?"; α=.93, M=3.3, SD=0.9). Interest in linear regression was measured post-instruction using four items (e.g., "How interesting do you find linear regression?"; α=.92, M=3.6, SD=1.2), and interest in the learning session was assessed with six items (e.g., "I've enjoyed this session."; α=.94, M=4.7, SD=1.0).

Confidence-related outcomes included a two-item expectancy measure for the upcoming test (e.g., "How well do you think you will do on the test?"; α=.92, M=3.1, SD=1.1) and a three-item measure of confidence in regression abilities (e.g., "How well do you think you would do in a regression course?"; α=.91, M=3.7, SD=0.9). Participants also provided a single-item judgment of learning ("The instruction I just received prepared me well to answer questions about linear regression."; M=4.6, SD=1.0). Measures were adapted from Linnenbrink-Garcia et al. (2010), Durik et al. (2015), Hecht et al. (2021), and Asher and Harackiewicz (2024), with judgments of learning drawn from Koriat and Ackerman (2010). Posttest performance was scored separately for memory items (M=38%, SD=37%) and generalization items (M=33%, SD=29%).

Results

We used preregistered linear regression models to test the effects of instructional condition on each outcome. Each outcome variable was regressed on two dummy-coded condition contrasts, using the explanatory feedback condition as the reference group. All models included standardized baseline math confidence and its interaction with each condition contrast. These interactions tested whether practice-centered instruction would be demotivating for less-confident students and allowed us to use confidence as a proxy for prior knowledge. We also fit exploratory models using lecture as the reference group to directly compare lecture to correct-response feedback. Descriptive statistics, correlations, and full regression output are reported in the Supplemental Materials.

Performance

Figure 2 shows average memory (3A) and generalization scores (3B), by condition.

As predicted, participants performed better on memory items when they practiced them and received feedback, compared to when they passively watched an instructor

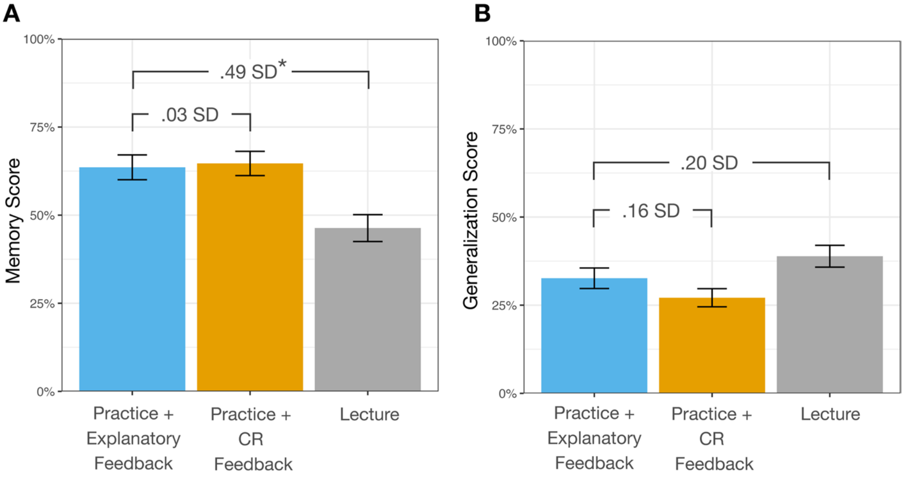

Fig. 2 Performance in Study 1: Memory and Generalization Scores. *Note.* CR = Correct Response. Error bars represent ±1 standard error of the mean (SEM). Effect sizes indicate standardized differences between each condition and practice with explanatory feedback. * An asterisk indicates a statistically significant difference, $p < .05$

complete the problems. Specifically, participants who received explanatory feedback outperformed those in the lecture condition on memory items by 17 percentage points (or 0.49 SD), $t(291) = -3.39$, $p = .001$. The advantage of correct-response feedback over lecture was similar: 18 percentage points, $t(291) = 3.56$, $p < .001$. There was no significant difference between the two practice conditions, $b = .01$, $t(291) = 0.15$, $p = .882$, suggesting that for memory, feedback type did not significantly impact performance. Interactions with baseline confidence were nonsignificant ($ps \geq .421$).

For generalization scores, neither lecture nor correct-response feedback differed significantly from explanatory feedback, though participants performed numerically better with explanatory feedback than correct-response feedback (33% vs. 27%, a.16 SD difference), $t(294) = -1.17$, $p = .245$, and numerically worse than lecture (33% vs. 39%, a.20 SD difference), $t(291) = 1.51$, $p = .131$. However, there was a significant feedback type × baseline confidence interaction, $b = -.08$, $t(291) = -2.22$, $p = .027$ (Fig. 3), indicating that whereas students in the correct-response condition performed poorly regardless of confidence, those with higher confidence benefited from explanatory feedback. For example, participants one standard deviation above the mean in baseline confidence scored 44% on generalization items, 13 points (or.45 SD) higher than their equally confident peers in the correct-response group, $t(291) = -2.35$, $p = .020$.

In the exploratory model comparing lecture to correct-response feedback, correct-response feedback was significantly less effective at promoting generalization, $t(291) = -2.69$, $p = .008$, a.36 SD difference (39% vs. 27%). This effect was also moderated by confidence, $b = -.09$, $t(291) = -2.41$, $p = .017$, suggesting that like explanatory feedback, lecture was ineffective at promoting generalization for less-confident students, Fig. 3.

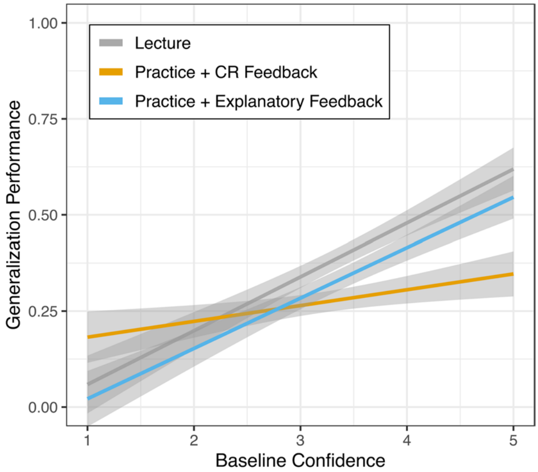

Fig. 3 Generalization Performance in Study 1: Condition × Confidence Interaction

Instruction Time

Participants in the explanatory feedback condition completed the learning session in an average of 7.5 minutes, compared to 9.9 minutes in the lecture condition, $t(291) = 4.00$, $p < .001$—a time savings of 2.4 minutes, or approximately 24%. In the correct feedback condition, participants completed the session in an average of 7.3 minutes—2.6 minutes faster than lecture, $t(291) = -4.20$, $p < .001$. There was no significant difference in time between the two practice conditions, $t(291) = -0.19$, $p = .848$, and interactions with baseline confidence were not significant ($ps \geq .101$). Given that, compared to lecture, participants in the elaborated feedback condition achieved much higher performance on memory questions, and roughly comparable performance on generalization questions, we can conclude that practice with elaborated feedback was more efficient.

Judgment of Learning and Confidence

On average, participants rated explanatory feedback as more effective preparation than correct-response feedback, $b = -0.38$, $t(291) = -3.00$, $p = .003$, a.37 SD difference. Judgments of learning were also more favorable for lecture compared to correct-response feedback, though not significantly, $b = -.22$, $t(291) = 1.78$, $p = .077$. There were no significant differences between lecture and explanatory feedback, $b = -.16$, $t(291) = -1.24$, $p = .217$, and no interactions with baseline confidence ($ps \geq .147$).

Confidence about the upcoming test and regression ability did not differ between conditions ($ps \geq .609$) or interact with baseline confidence ($ps \geq .576$).

Interest in the Learning Session and Linear Regression

Self-reported interest in regression did not differ between conditions ($ps \geq .545$) or interact with baseline confidence ($ps \geq .136$). Situational interest in the learning session also did not differ across conditions (*p*s > .655). However, a nonsignificant trend suggested that lecture may have been more engaging than explanatory feedback for less-confident students, $b = -0.21$, $t(291) = -1.69$, $p = .092$, consistent with past findings (Asher, Sana, et al., 2025).

Discussion

The results of Study 1 suggest that students can effectively learn from practice with feedback, but memory and generalization likely require different instructional supports. Practice with feedback was clearly more effective than lecture for promoting memory: participants in both feedback conditions outperformed those in the lecture condition by nearly half a standard deviation, with a 24% time savings.

For generalization, the pattern was more nuanced. Correct-answer feedback was insufficient: participants in this group performed poorly regardless of initial confidence. In contrast, students who received either lecture or elaborated feedback performed significantly better—but only if they entered the study with high math confidence. These findings indicate that conceptual instruction—whether delivered via lecture or embedded in elaborated feedback—is necessary but may not be sufficient for generalization unless learners are ready to engage with complex explanations.

Study 1 provided no clear evidence that practice with feedback undermines motivation. Participants in the explanatory feedback condition rated their instruction as more effective than those who received only correct-response feedback, but neither approach significantly affected confidence relative to lecture. Interest in the topic and learning session did not differ significantly across conditions, though there was a nonsignificant trend suggesting that lectures may have been more engaging than explanatory feedback for less-confident students.

Despite these promising findings, Study 1 had several limitations. First, because we omitted a regression pretest, we cannot determine whether confident students benefited more from conceptual instruction (lecture or explanatory feedback) due to cognitive readiness (prior knowledge) or greater willingness to engage with challenging material (motivational readiness). Second, the multiple-choice format for both practice and memory test questions may have inflated memory scores by encouraging item-specific recognition or reliance on superficial features rather than genuine recall (see Glover, 1989; Kang et al., 2007; Roediger & Karpicke, 2006). Although factual recall is important in many domains, it is essential to ensure that learning from practice extends beyond answer recognition to knowledge that students can genuinely recall and apply. Third, our measure of judgments of learning assessed perceived instructional effectiveness rather than expected performance, preventing us from assessing metacognitive calibration—a key component of our theoretical framework about how practice with feedback supports learning.

To replicate and build upon Study 1, addressing its limitations, we conducted Study 2.

Study 2

Study 2 replicated and extended Study 1 with three main goals. First, to disentangle confidence from prior knowledge, we introduced a brief pretest of statistical reasoning to directly measure participants' baseline understanding. Second, to assess whether practice with feedback improves metacognitive calibration, we asked participants to predict their test performance before completing the posttest, allowing us to evaluate the alignment between their expectations and actual outcomes. Third, to address the limitations of the multiple-choice format used in Study 1, we redesigned both the practice session and posttest to require exclusively open-ended responses. We expected that this shift would clarify whether practice-centered instruction supports meaningful recall rather than superficial recognition and might enhance instructional value, as generating answers requires greater cognitive effort than recognizing correct choices (see Kang et al., 2007).

However, the shift to open-ended responses introduced a technical challenge: while multiple-choice items are easy to grade automatically, responses with open-ended reasoning are more difficult to score and respond to in real time. In prior research on retrieval practice, open-ended feedback has typically consisted of simply showing the correct answer after each response (e.g., Butler et al., 2013; Kang et al., 2007). Recent advances in generative AI offer a more interactive alternative. Large language models such as GPT can now evaluate open-ended responses and generate personalized explanations about why answers are correct or incorrect. In Study 2, we used GPT-4o to provide immediate, tailored explanatory feedback after each practice item, enabling interactive, personalized feedback with free-response practice questions.

Hypotheses

We again predicted that practice with either type of feedback would be more effective than lecture at promoting memory (H1), but that explanations would be needed for generalization (H2). For generalization we also predicted that explanations might be unproductive for participants with low prior knowledge, consistent with cognitive load theory (H3). We predicted that both forms of practice would be more efficient than lecture (H4), and that participants with less confidence might be discouraged by the difficulty of solving practice problems without upfront guidance, demonstrating decreased interest in the learning session (H5). Finally, we predicted that practice with either type of feedback would decrease overconfidence, leading to better calibration between participants' predicted and actual test scores (H6).

This study was approved by the Institutional Review Board at Carnegie Mellon University. Hypotheses, along with data collection and analysis procedures were preregistered at https://osf.io/28jtr.

Method

Participants

Participants were recruited online via *Prolific*, and they were required to be at least 18 years old and living in the United States. Following our preregistration, data collection continued until 300 participants consented and met all inclusion criteria. For Study 2, we also implemented checks to detect use of generative AI on the posttest. Participants flagged for AI use were excluded and replaced; in total, 41 participants (12% of those who completed the study) were excluded. Demographic data were not available for excluded participants.

In the final sample of 300 participants who met inclusion criteria, 70% self-reported their ethnicity as White, 20% as Black, 4% as multiracial, 3% as Asian, and 2% as belonging to another group; 1% did not report their ethnicity. Half identified as women and half as men. The average age was 39.2 years. In terms of math background, 47% reported completing a college-level math course, 17% had completed an advanced placement math course, and 34% had no math coursework beyond a standard high school level. The study lasted approximately 30 minutes, and participants were paid $6.00.

Procedure

Study 2 was administered via *Qualtrics* using the same three-cell design as Study 1: participants learned about multiple regression with either (1) lecture, (2) practice with correct-response feedback, or (3) practice with explanatory feedback. Participants completed a baseline questionnaire, a statistics lesson, an outcome questionnaire, and a posttest. However, we introduced several changes to simplify the lesson, better measure prior knowledge, and improve the instructional design.

As in Study 1, participants were told they would be learning about linear regression and then completing a test. The baseline questionnaire included the same three-item measure of math confidence, and we added a four-question pretest of general statistical reasoning to directly measure prior knowledge (see Measures below).

To simplify the learning session, we narrowed the content from six to three learning objectives: identifying the intercept of a line and interpreting the slope and intercept. In the lecture condition, this reduced the video to approximately 7.5 minutes. The practice-based conditions were shortened to three practice problems matching the three targeted objectives.

We also simplified the posttest to mirror the format of the lecture and practice. Participants answered six questions: three “memory” questions that exactly matched the three questions presented in the practice or lecture, and three “generalization” questions that required applying the same skills with a new line. By testing the same learning objectives in the same format on both practice and posttest, this allowed us to measure efficiency (change in knowledge over time; Koedinger et al., 2012) more precisely in the two practice conditions by using practice scores as estimates of initial knowledge.

The most substantial change was the shift from multiple-choice to open-ended responses for both practice and testing. In the correct-response feedback condition, participants were told whether their answer was correct and shown the correct response. In the explanatory feedback condition, responses were evaluated using GPT-4o, which provided instant, personalized explanatory feedback. This feedback identified whether the answer was correct and offered an AI-generated rationale for incorrect responses, and it also included the same standardized visuals used in Study

1 (see Fig. 4 for an example). To provide AI feedback through *Qualtrics*, we adapted a procedure used by Costello et al. (2024).

Table 2 displays prompts used to provide feedback for participants' interpretation of the line's intercept. For this question and all others, GPT-4o was provided with a written explanation of the problem and told to provide feedback consisting of "Correct!" or "Incorrect." For incorrect answers, the prompt asked GPT to explain why answers were incorrect and to "edit their exact wording with the smallest possible changes to make their response fully correct," labeling this the "corrected version." GPT was given the specific criteria expected in a correct answer and sample correct and incorrect responses based on pilot work to ensure feedback accuracy across a wide range of responses.

Three sample incorrect responses from participants in Study 2 are displayed in Table 3, along with the corresponding AI feedback for each response.

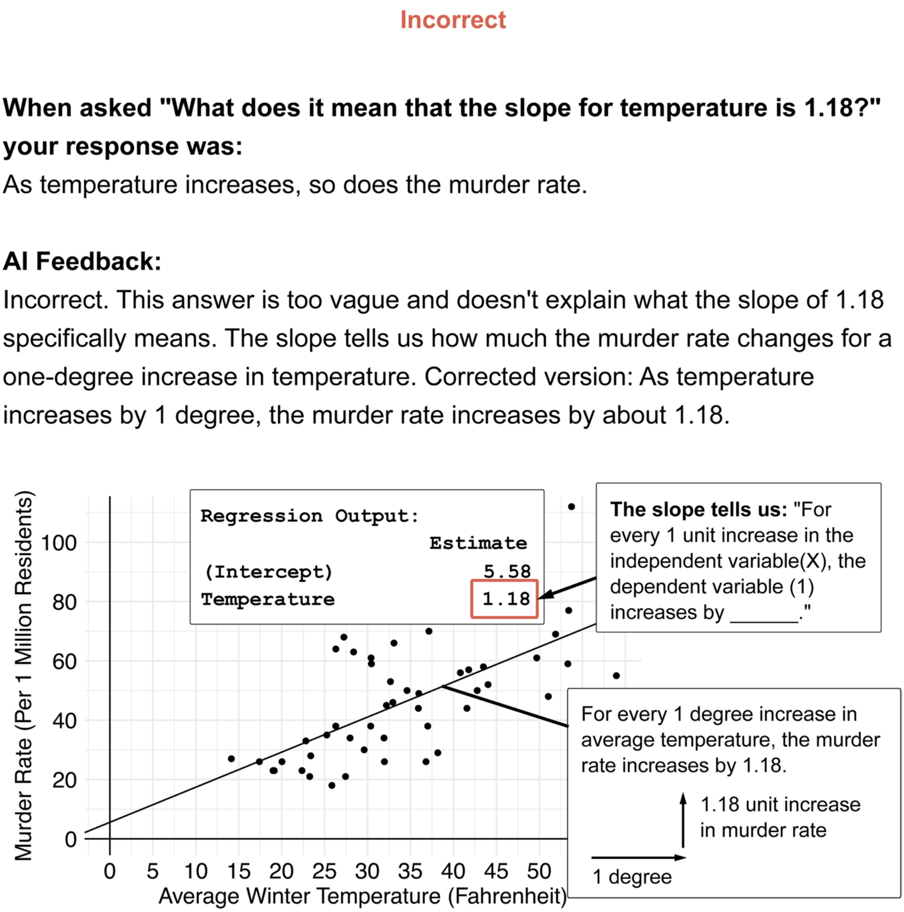

Fig. 4 Sample AI-Assisted Explanatory Feedback, Study 2. *Note.* The "AI Feedback" message was generated by GPT-4o in response to the student's answer

Table 2 The Prompt Given to GPT-4o to Elicit Feedback about Interpreting an Intercept

A student was shown a scatterplot with the description, "In the scatterplot below, each point represents a U.S. state. "The plot shows the relationship between each state's average (winter) temperature and its murder rate. A regression line, determined by a computer, is shown on the plot." They were told that the slope of the regression line was 1.18, and that the intercept was 5.58. They were then asked: "What does it mean that the intercept is 5.58?" Give them feedback about the correctness of their answer. Your feedback should contain only the following: (1) The word "Correct!" or "Incorrect." (2) if the answer is wrong, explain exactly why. (3) if the answer is wrong, then say "Corrected version:" and edit their exact wording with the smallest possible changes to make their response fully correct. A correct answer should communicate that the intercept tells us the predicted murders (or murder rate) when temperature equals zero. They don't need to say 5.58 in their response (and it's okay if they round the number to 6, or say something like "between 5 and 6″), but they do need to say something about temperature equaling zero. Address the student directly. Here is the student's response: [RESPONSE HERE]

Table 3 Sample Incorrect Interpretations for an Intercept, along with Corresponding AI Feedback

Incorrect Response	AI Feedback
It intersects the Y axis at 5.58	Incorrect. The intercept does occur at 5.58 on the y-axis, but this explanation doesn't tell us what that means in the context of the data. Corrected version: It means that when the average winter temperature is zero degrees, the predicted murder rate is 5.58.
Change in y over change in the value of the x and y coordinate	Incorrect. The intercept is not a description of the change in values — that describes the slope. The intercept represents the predicted value of the y-variable (murder rate) when the x-variable (temperature) is zero. Corrected version: The intercept tells us the predicted murder rate when the temperature is zero.
Thats where the line crosses the y intercept and it means the murder rate it 5.58 when the temps are at 1.18	Incorrect. The intercept tells us the predicted value of the murder rate when the temperature is **zero**, not when it's 1.18. 1.18 is the slope, not the temperature value at the intercept. Corrected version: "That's where the line crosses the y-axis, and it means the murder rate is 5.58 when the temperature is zero."

To validate the AI feedback, we manually reviewed all 309 AI-generated responses and determined that the AI correctly scored 90% of correct answers and 98% of incorrect answers, and when providing corrections, it offered appropriate feedback 98% of the time with no hallucinations (all feedback was grounded in participants' actual responses and our scoring criteria). For example, the prompt in Table 2 instructed the AI that a correct interpretation of an intercept "should communicate that the intercept tells us the predicted murders (or murder rate) when temperature equals zero." When a student wrote only "It intersects the Y axis at 5.58," the AI correctly identified the missing interpretation and provided a valid corrected answer: "The intercept does occur at 5.58 on the y-axis, but this explanation doesn't tell us what that means in the context of the data … It means that when the average winter temperature is zero degrees, the predicted murder rate is 5.58." Overall, the AI provided reliable formative feedback that was sufficiently accurate for our research purposes. Details about the evaluation process are provided in the Supplemental Materials, along with prompts, sample responses, and feedback for all three practice questions.

Following the learning session, participants completed an outcome questionnaire measuring interest and judgments of learning. Immediately before the posttest, participants predicted their score on the upcoming test, allowing us to measure metacognitive calibration. The posttest consisted of six open-ended questions: three memory items aligned with content from the lesson and three generalization items requiring application of the same concepts to new problems.

Measures

To measure participants' background knowledge of statistics, we selected four questions from a pilot version of the statistical Reasoning and Literacy (REALI) examination (Sabbag, 2016), using item response theory statistics to select items with a range of difficulty and high discrimination. These items test knowledge of statistics terms and concepts (sample vs. population, null hypothesis testing, confidence intervals; $M=39\%$, $SD=31\%$). Internal consistency for the selected items was modest ($\alpha=.58$), which is expected given the scale's brevity and the fact that each item assessed a distinct statistical concept (Cortina, 1993). However, the measure's correlation with posttest scores ($r=.34$) suggests that it captures meaningful variation in general statistics knowledge.

Measures for math confidence ($M=3.1$, $SD=1.0$, $\alpha=.92$), judgments of learning ($M=4.4$, SD$=1.4$), situational interest ($M=4.5$, SD$=1.2$, $\alpha=.95$), and interest in regression ($M=3.4$, SD$=1.1$, $\alpha=.95$) were identical to those used in Study 1. Posttest performance was scored separately for memory ($M=44\%$, $SD=36\%$) and generalization items ($M=35\%$, $SD=35\%$), using a rubric shared in the Supplemental Materials.

For a measure of metacognitive calibration, participants predicted the percentage of posttest questions they would answer correctly immediately before taking the posttest ($M=52.5\%$, $SD=26.9\%$; see Hartwig & Dunlosky, 2017). We calculated calibration accuracy as the difference between actual and predicted, with higher scores indicating greater overconfidence ($M=16.4\%$, $SD=36.5\%$).

Finally, to estimate the average efficiency of learning in each condition, we adopted an imputation procedure used by Asher, Sana, et al. (2025). We estimated initial knowledge for participants in the practice-based conditions using their practice session performance (M$=19.3\%$, SD$=29.0\%$). To estimate initial knowledge for participants in the lecture condition (who lacked practice scores), we used a resampling procedure to generate 1000 imputed datasets, each with different plausible estimates for students' initial knowledge. We then calculated efficiency scores for participants in each imputed dataset and analyzed the datasets separately, pooling regression estimates to yield an unbiased estimate of efficiency while reflecting the uncertainty in participants' true initial knowledge. This process is detailed in the Supplemental Materials.

Results

As in Study 1, we analyzed each outcome using preregistered linear regression models. Each model included two dummy-coded condition contrasts using the explanatory feedback condition as the reference group. For performance outcomes, models also included participants' standardized pretest scores and the interaction between pretest and condition, allowing us to test whether prior knowledge moderated instructional effectiveness. For motivation-related outcomes, math confidence was included instead of pretest scores, consistent with Study 1, because we expected that participants' subjective beliefs about their ability would shape whether they appreciated the challenge of learning from practice and feedback.

Performance

Figure 5 shows average memory (5A) and generalization scores (5B), by condition. A dashed line at 19% serves as an estimate of participants' baseline performance before instruction, generated by calculating average performance during the practice session, which did not vary by condition (correct-response feedback $M=21\%$, elaborated feedback $M=18\%$).

Memory Participants in the explanatory feedback condition outperformed those in the lecture condition on memory questions by 13 points, $t(294)=-2.64$, $p=.009$, or 0.35 SD (Fig. 5A). Correct-response feedback was also better for promoting mem-

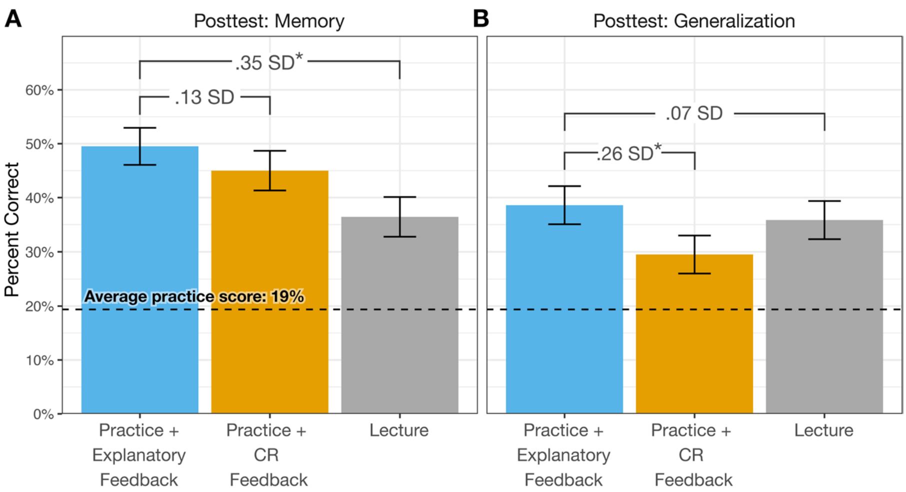

Fig. 5 Performance in Study 2: Memory and Generalization Scores. *Note.* CR = Correct Response. Error bars represent ±1 SEM. The dashed line shows average practice session performance (an initial knowledge estimate). Effect sizes indicate standardized differences between each condition and practice with explanatory feedback. * An asterisk indicates a statistically significant difference, $p<.05$

ory than lecture (by 8 points), but this difference was not statistically significant, $t(294)=1.65$, $p=.099$. As was the case in Study 1, there was no significant difference between the two practice conditions, (50% vs. 45%, $t(294)=-0.99$, $p=.325$). Interactions between condition contrasts and pretest scores were not statistically significant (ps$\geq$.475).

Generalization Participants who received explanatory feedback scored nine points higher on generalization than those in the correct-response feedback condition, $t(294)=1.97$, $p=.050$, a difference of.26 SD. There was no significant difference between explanatory feedback and lecture (36% vs. 39%, $t(294)=-0.51$, $p=.609$). Although the Correct-Response Feedback × Pretest interaction was not significant, $b=-0.06$, $t(294)=-1.28$, $p=.202$, the effect was in the predicted direction, Fig. 6A.

In the secondary model comparing lecture directly to correct-response feedback, correct-response feedback was found to be 7 points less effective at promoting generalization than lecture, although this effect was not significant in Study 2, $t(294)$, $p=.153$. In addition, the Lecture × Pretest interaction was significant, $b=.11$, $t(294)=-2.35$, $p=.019$, further suggesting that participants with higher prior knowledge benefited more from instruction that included conceptual explanations provided in the lecture, compared to correct-response feedback (Fig. 6A).

Efficiency of Instruction

Figure 7 shows the estimated average efficiency of learning for memory (7A) and generalization (7B) in each condition. For memory scores, participants in the practice with elaborated feedback condition learned at an estimated rate of 6.93 points per minute, 3.4 times as efficiently as those in the lecture condition (2.05 points per minute), $t(265.8)=-4.22$, $p<.001$. Participants in the practice with correct-response feedback condition learned at 5.32 points per minute, 2.6 times as efficiently as lec-

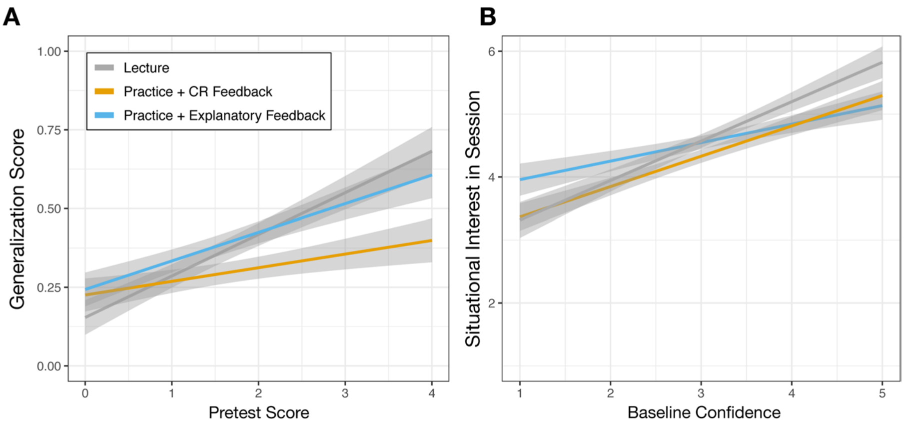

Fig. 6 Interactions in Study 2 with Pretest Score and Confidence. *Note.* CR = Correct Response. Error bars represent ± 1 standard error of the estimate

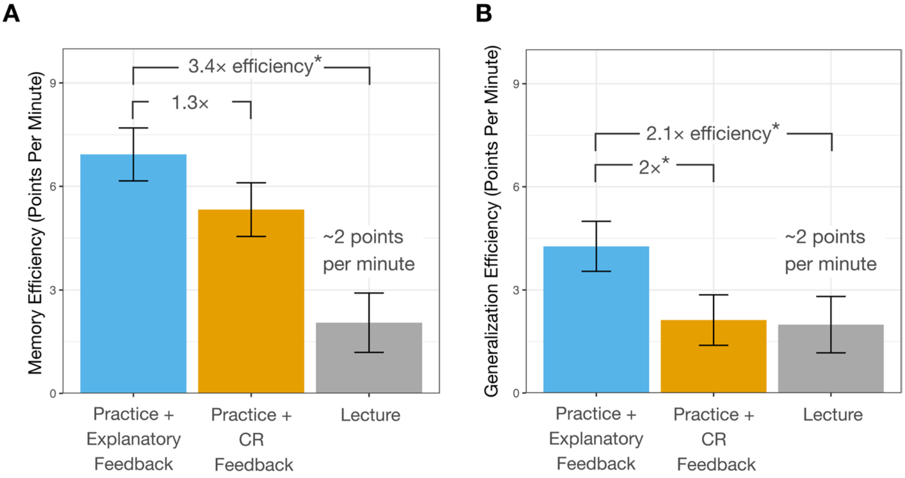

Fig. 7 Average Learning Efficiency in Study 2 by Condition on Memory and Generalization Questions, *Note*. Figure 7 displays predicted values from the model we fit to estimate efficiency in each condition; error bars represent ±1 standard error of each estimate. * An asterisk indicates a statistically significant difference, $p<.05$

ture, $t(266.2)=2.78$, $p=.006$. Memory efficiency estimates did not significantly differ between the two practice conditions, $t(292.0)=-1.50$, $p=.134$. There were no significant interactions with pretest scores, $ps \geq .224$.

For generalization scores, participants in the elaborated feedback condition learned at an estimated rate of 4.27 points per minute, 2.1 times as efficiently as those in the lecture condition (1.99 points per minute), $t(262.9)=-2.06$, $p=.041$, and twice as efficiently as those who learned from correct-response feedback (2.12 points per minute), t(262.9) = −2.06, $p=.036$. For generalization, there was no significant difference in efficiency between correct-response feedback and lecture, $t(263.3)=0.07$, $p=.942$. There were no significant interactions with pretest scores, $ps \geq .200$.

Metacognitive Monitoring

Judgment of Learning In all conditions, participants tended to agree that their instruction was effective at preparing them to answer questions about linear regression, with average responses in the "agree" to "strongly agree" range, Fig. 8A. There were no significant differences in judgments of learning between the lecture and explanatory feedback conditions, $d=.04$, $t(294)=0.24$, $p=.810$. However, practice with explanatory feedback was judged to be more effective than practice with correct-response feedback, $d=-.39$, $t(294)=-2.24$, $p=.026$, as was lecture, $d=-.43$, $t(294)=-2.45$, $p=.015$, replicating Study 1. No condition × baseline confidence interactions were significant ($ps \geq .137$).

Metacognitive Calibration As predicted, participants in the lecture condition were significantly more overconfident, Fig. 8B. On average, they overestimated their test performance by 26 percentage points. In contrast, participants in both practice condi-

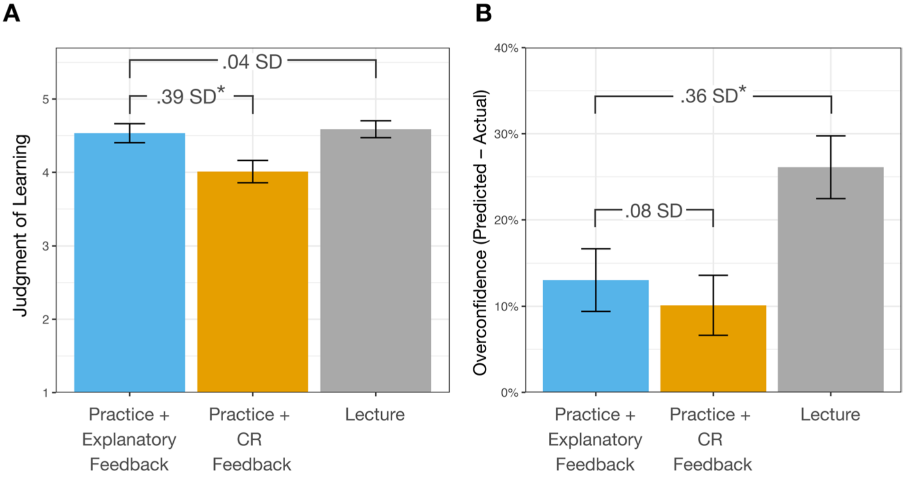

Fig. 8 Judgment of Learning and Overconfidence (a Measure of Metacognitive Calibration), Study 2. *Note.* CR = Correct Response. Error bars represent ±1 SEM. Effect sizes indicate standardized differences between each condition and practice with explanatory feedback. * An asterisk indicates a statistically significant difference, $p<.05$

tions were better calibrated. Compared to lecture, average prediction errors were 13 points lower in the explanatory feedback condition, $t(294)=2.69$, $p=.008$, and 13.6 points lower in the correct-response feedback condition, $t(294)=2.77$, $p=.006$. The two practice conditions did not differ significantly from each other: participants overestimated their performance by 13 points with explanatory feedback and 10 points with correct-response feedback, $t(294)=0.11$, $p=.912$. No condition × baseline confidence interactions were significant ($ps \geq .449$).

Interest in the Learning Session and Linear Regression

There were no significant differences between conditions in participants' interest in regression or the learning session, $ps \geq .174$. Likewise, no significant interactions with math confidence emerged, except for one unexpected finding: a positive *Lecture* × baseline confidence interaction for interest in the learning session, $b=0.32$, $t(294)=2.03$, $p=.043$. In contrast to Study 1, this interaction showed that explanatory feedback was rated as more engaging than lecture by participants with lower math confidence, whereas lecture was more engaging for those with higher confidence (see Fig. 6B).

Discussion

Study 2 replicated and extended Study 1, reinforcing that students can efficiently learn through practice with feedback and demonstrating that these benefits persist even when both practice and test items require open-ended responses. By moving beyond multiple-choice formats, Study 2 showed that practice-centered instruction can promote meaningful recall and generalization, not merely recognition of answer choices.

As in Study 1, correct-answer feedback was sufficient to support memory, but explanations remained necessary to promote generalization. Critically, Study 2 provided clearer evidence that prior knowledge—directly measured through a statistics pretest—moderated these effects: learners with higher pretest scores benefitted more from elaborated explanations, whereas those with lower pretest scores gained little from either lecture or elaborated feedback. This finding clarifies that foundational understanding may be a prerequisite for instruction to effectively support generalization, consistent with cognitive load theory. An efficiency analysis further highlighted practical advantages: for generalization, elaborated feedback yielded over 2x greater learning per minute than lecture; for memory, both practice conditions produced more than 2.5x learning per minute compared to lecture.

Study 2 also extended Study 1 by examining metacognitive calibration. Practice, whether paired with simple or elaborated feedback, significantly reduced overconfidence compared to lecture. Participants who only watched a lecture overestimated their performance by an average of 26 percentage points, whereas those who practiced and received feedback improved their calibration accuracy by more than 50%. This pattern suggests that practice-based instruction provides metacognitively useful opportunities for self-assessment during problem-solving.

Explanatory feedback was again judged to be more effective than practice with correct-response feedback, but similarly effective to lecture, replicating Study 1, with no differences in confidence across conditions. This pattern suggests learners appreciate conceptual explanation—whether delivered via lecture or feedback—but struggle to calibrate how much they have learned without also practicing and seeing evidence about their performance.

Finally, there were no overall effects on participants' interest, although an unexpected interaction revealed that explanatory feedback was more engaging than lecture for participants with lower math confidence—reversing the trend observed in Study 1.

General Discussion

Across two studies, we investigated when practice-centered instruction can succeed, as well as the affordances of this instructional strategy for memory, generalization, metacognition, and motivation. Building on recent demonstrations that practice with feedback has the potential to replace lecture (Asher, Sana, et al., 2025), we experimentally manipulated feedback quality to test whether it matters for different learning outcomes. Our results replicate and extend this prior work, demonstrating for the first time that students can both memorize and learn to generalize from practice alone, but only when feedback is appropriately matched to the learning objective and students possess sufficient prior knowledge. Critically, we found that prior knowledge is equally important for learning from video lectures, which proved comparably effective but less efficient for generalization. In addition, practice with elaborated feedback offered advantages for memory and metacognitive calibration, without detectable negative consequences for motivation.

Boundary Conditions: When Practice-Centered Instruction Will Succeed

Studies 1 and 2 demonstrate that the effectiveness of practice with feedback depends on two critical boundary conditions: (1) the type of learning objective—memorization versus generalization—and (2) the content of feedback—correct answers versus explanations of underlying principles. These boundary conditions interact to determine when practice with feedback can successfully replace lecture.

For memory outcomes, practice with feedback consistently enhanced retention, regardless of feedback type. In both studies, participants who engaged in practice with feedback—whether simple or explanatory—outperformed those who received lecture-based instruction on recognition tests (Study 1) and recall tests (Study 2). This indicates that feedback content matters less when the learning goal is memorization, and any practice with feedback will match or beat a good lecture.

For generalization outcomes, practice with feedback reached the level of a good lecture only when elaborated feedback was presented. Correct-answer feedback was insufficient to promote transfer to novel problems, demonstrating that feedback content is decisive when the learning goal involves forming new generalizations. Students must be told not just what the right answer is, but why it is correct. Explanatory feedback—which offered underlying rules or principles—consistently outperformed correct-answer feedback for generalization across both studies.

Ingredients for Promoting Memory Versus Generalization

Within these boundary conditions, specific ingredients enable practice with feedback to support different learning outcomes. The ingredients required differ substantially depending on whether the goal is memory or generalization.

For memory outcomes, the essential ingredients are straightforward: practice attempts paired with corrective feedback containing the correct answer. Both studies demonstrated that this minimal combination is sufficient to support retention. The practice attempt engages active processing and metacognitive calibration, while feedback—even when it simply indicates the correct answer—provides the information needed to correct errors and consolidate accurate knowledge. No additional elaboration or explanation is necessary when students need only to remember associations.

Although a lecture can promote memory if learners attend carefully, its passive nature makes it vulnerable to distraction (Chi & Wylie, 2014). In contrast, practice offers a built-in metacognitive check that can support encoding: attempting an answer, then receiving any type of corrective feedback, naturally highlights gaps in knowledge and engages deeper processing as students seek relevant information (see also Carvalho et al., 2024).

Studies on test-enhanced learning also suggest that if after encoding an answer, learners have the opportunity to continue practicing and retrieving from memory, retrieval practice should decrease forgetting and improve longer-term recall (see Pan & Carpenter, 2023; Roediger & Karpicke, 2006). However, in Studies 1 and 2 participants were not given repeated practice or delayed posttests, so retrieval practice is an unlikely mechanism to explain the memory benefits observed.

Supporting generalization is more complex, with two necessary ingredients: explanatory feedback and sufficient prior knowledge.

First, explanatory feedback is necessary. Unlike memory, which can be supported by simple corrective feedback, generalization requires feedback that explains underlying principles and rules. This explanatory content functions similarly to lecture in providing the conceptual foundation needed for transfer (Sweller, 2004), but delivers it in a more targeted and timely manner—immediately following a student's attempt to apply the concept.

Second, learners must possess sufficient prior knowledge to engage productively with explanatory feedback. This ingredient emerged clearly across both studies. In Study 1—where we lacked a pretest of prior knowledge—learners with higher math confidence generalized well from elaborated feedback. Study 2, which contained a pretest, clarified that generalization depended directly on prior knowledge. These findings support cognitive load theory (Kirschner et al., 2006), which posits that learners need sufficient background knowledge to process complex explanations productively—whether those explanations are delivered through lecture or feedback. When cognitive load exceeds learners' capacity, even high-quality explanations may fail to support transfer. Critically, this pattern held for both lecture and practice with elaborated feedback in our studies, suggesting that prior knowledge is a general prerequisite for learning to generalize rather than a unique requirement of practice-centered instruction.

When (and When Not) to Use Practice with Feedback Instead of Lecture

The decision to use practice with feedback instead of lecture should be guided by alignment with these boundary conditions and availability of the necessary ingredients.

Practice with feedback is highly advantageous when time efficiency is critical. In Study 2, elaborated feedback yielded over 2x greater learning per minute for generalization and more than 3x for memory compared to lecture. Importantly, while practice with feedback proved substantially more time-efficient than lecture, lecture was no less effective in terms of overall learning when time was not constrained. In both studies, lecture and elaborated feedback produced similar generalization outcomes, with lecture even showing a non-significant advantage compared to multiple-choice practice in Study 1. This suggests that when time is not a limiting factor, lecture retains instructional value for building conceptual understanding.

Practice with feedback is also advantageous when learning goals align with available feedback types: memorization (which requires only simple feedback) or generalization for prepared learners (which requires explanatory feedback and adequate prior knowledge). The importance of prior knowledge as an ingredient suggests that explanatory feedback is best suited for learners with at least moderate preparation in the domain, or for instructional sequences where initial practice sessions focus on building foundational knowledge through carefully scaffolded problems before advancing to more complex applications.

Conversely, practice-only instruction may be less effective than lecture when it is easier to tailor lecture to diverse preparation levels—for instance, in small classroom settings where instructors can adjust explanations in real-time based on student questions and visible confusion. Additionally, some topics might require extended explanation of highly abstract principles that resist decomposition into brief, targeted feedback. Although we have not identified such cases in our studies, if they exist, lecture may offer value for building conceptual understanding by providing comprehensive explanations that brief feedback cannot deliver.

Metacognitive and Motivational Outcomes of Practice with Feedback

Beyond cognitive outcomes, our studies examined how practice with feedback affects learners' metacognitive awareness and motivation compared to lecture. The findings reveal consistent metacognitive benefits and mixed but generally neutral motivational effects.

Benefits for Metacognitive Calibration

In Study 2, students who practiced—regardless of feedback type—demonstrated substantially better calibration between their predicted and actual performance than those who learned by lecture. Lecture participants overestimated their performance by an average of 26 percentage points, a pattern consistent with research showing that passive instruction often leads to overconfidence (Bjork et al., 2013). In contrast, attempting problems and receiving immediate feedback gave practice-based learners clearer insight into their knowledge gaps, improving calibration accuracy by more than 50%.

This improvement occurred without additional scaffolding such as reflection prompts, indicating that the structure of practice with feedback inherently supports metacognitive development. The act of attempting an answer forces a moment of self-assessment—students must evaluate whether they can solve the problem—and feedback provides immediate validation or correction. These benefits suggest that practice with feedback not only supports memory and generalization but also offers critical cues for more accurate self-evaluation, which may support subsequent self-regulated learning.

Motivational Outcomes

Despite concerns that feedback-based instruction might be discouraging—especially for learners with low confidence—we found no consistent negative effects on learners' interest or confidence across studies. In both studies, participants reported similar levels of confidence after all types of instruction. Regarding judgments of instructional effectiveness, practice with correct-response feedback was judged to be less effective than both lecture and elaborated feedback, but elaborated feedback was rated as similarly effective to lecture. This pattern suggests that learners appreciate conceptual explanation, whether delivered through lecture or feedback, but view simple corrective feedback as less valuable—even though it still supports memory.

Interest outcomes were more complex and varied between studies. In Study 1, lecture produced somewhat greater situational interest than elaborated feedback for less-confident learners, supporting concerns that practice-first approaches might feel discouraging to underprepared students. However, Study 2 revealed a significant interaction in the opposite direction: elaborated feedback—delivered through open-ended practice with personalized, AI-powered explanations—elicited higher interest among less-confident learners compared to lecture.

These discrepant findings may reflect differences in practice format (multiple choice versus open ended) or feedback delivery (generic versus personalized), or they may simply represent the instability of interactions involving continuous moderators, which are often difficult to detect and replicate (Asher, Hecht, et al., 2025; McClelland & Judd, 1993). Taken together, however, the findings suggest that practice with feedback may not systematically harm motivation relative to lecture, and may even enhance engagement under certain conditions. Future researchers should investigate questions about moderation with larger sample sizes and head-to-head comparisons of different types of explanatory feedback to better understand when practice-based instruction can be motivating or demotivating.

Overall, these findings indicate that practice with feedback offers clear metacognitive advantages over lecture without consistent motivational costs. The improved calibration that practice affords may be particularly valuable for supporting self-regulated learning, as accurate self-assessment is essential for students to identify what they need to study and how to allocate their learning efforts effectively.

Implications for Instruction with Generative AI

In addition to clarifying the affordances and mechanisms of practice with feedback, Study 2 demonstrates the potential for generative AI to expand the scope of feedback-based instruction. By using GPT to evaluate open-ended responses and generate personalized explanatory feedback, we enabled real-time scoring and feedback for this cognitively demanding response format. This approach offers a scalable solution for providing responsive, explanatory feedback in contexts like large-enrollment courses where immediate human evaluation is impractical. As large language models become increasingly integrated into educational technologies, this work demonstrates how they can be used not just to assess responses, but to support learning.

Limitations and Future Directions

This research was designed to test the boundary conditions under which practice with feedback can support memory and generalization without upfront lecture, using tightly controlled experimental manipulations within brief online sessions. While our findings clarify when and for whom this instructional approach can succeed (Mook, 1983), this design choice necessarily involved trade-offs that point toward important directions for future work.

First, both studies involved brief instructional sessions (under 10 minutes) with immediate posttests. This brevity enabled precise experimental control but limits ecological validity. Authentic classroom instruction unfolds over longer periods, covering more complex content with repeated practice opportunities. Moreover, our outcome measures used researcher-developed tests closely aligned with instructional content rather than standardized achievement measures. While this alignment was intentional—allowing us to isolate effects of our manipulations—it may overestimate practical benefits (Bardach et al., 2025). For instance, generalization items required applying learned principles to novel problems with new data, but maintained the same question formats participants encountered during instruction.

Replication in authentic educational settings is essential to assess whether practice-centered learning can build knowledge of complex topics over time and improve more distal outcomes such as course grades or standardized assessments. In this work, researchers should examine optimal methods for replacing lectures with practice and feedback, which will vary substantially between contexts. For example, teachers may provide personalized feedback in small classrooms, while large courses might require instructor-curated feedback libraries or AI-assisted systems. Based on the boundary conditions we identified, these approaches may struggle when feedback cannot be sufficiently tailored to diverse preparation levels or when content requires extended explanation of highly abstract principles.

Second, our measures of prior knowledge were imperfect. Study 1 relied on math confidence as a proxy, while Study 2 used a four-item pretest with poor internal consistency (α=.58). Although both measures revealed the predicted moderating pattern, lower measurement reliability attenuates effect sizes and reduces statistical power (Bohrnstedt & Marwell, 1978); McClelland & Judd, 1993), meaning the observed moderation effects may be conservative. More comprehensive assessments are needed to consistently detect interactions between instructional strategies and prior knowledge, and to better understand the threshold at which learners can productively engage with explanatory feedback. Longer classroom studies would enable thorough pretesting, and measurement strategies that reduce error, like latent interaction modeling (Klein & Moosbrugger, 2000), could help maintain statistical power.

Third, our design included limited measurement of the psychological mechanisms proposed in our theoretical framework. Although Study 2 documented benefits for metacognitive calibration, we did not assess attention during instruction, cognitive load during feedback processing, or students' learning goals. Testing these mediators would strengthen causal claims about how practice with feedback supports learning. Future experiments could incorporate think-aloud protocols, self-reports of mental effort, and eye tracking during learning sessions.

Moreover, although we discussed memory, generalization, metacognitive monitoring, and situational interest as separate outcomes, practice with feedback may be particularly effective because it causes these processes to influence each other. We suggest that practice-based learning creates a psychological cascade: attempting problems triggers metacognitive monitoring, which reveals knowledge gaps and activates learning goals, which motivates attention toward feedback and supports encoding. Testing this cascade would require measuring multiple processes simultaneously and examining their temporal dynamics.

Fourth, while Study 2 provided evidence that GPT-4o can provide accurate and helpful feedback, our validation was limited to a single topic and one session. The consistency and quality of AI-generated feedback across diverse content areas and extended learning sequences remains unknown. Research is needed to understand how students perceive and engage with AI-generated versus human-generated feedback, and whether trust and acceptance vary across learner characteristics.

Finally, our findings suggest that practice with explanatory feedback is most effective when explanations are matched to learners' background knowledge. It is important to note that this does not rule out practice-centered instruction for underprepared learners; instead it suggests a pathway for supporting them. Rather than viewing low prior knowledge as a barrier requiring upfront lecture, initial practice sessions could build foundational understanding through carefully scaffolded problems and feedback. Future research should test specific scaffolding strategies—such as highly structured problems that break tasks into smaller steps, or prompting self-explanation during feedback—to identify which approaches can reduce cognitive load while preserving the metacognitive and motivational advantages of active engagement. Investigating these design variations, alongside deeper measurement of proposed mediators and moderators and longitudinal assessment of authentic educational outcomes, would advance both theory and practice.

Conclusion

This research advances a theory of practice-based instruction in which practice with feedback—when sufficiently elaborated and matched to learner's background knowledge—is not a *complement* to prior instruction but the instruction itself. As educational systems seek more efficient, scalable ways to personalize instruction, practice-based learning with elaborated feedback—delivered by humans or AI—may represent a powerful and flexible alternative to traditional lecture-based approaches.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10648-025-10103-6.

Author Contributions Michael W. Asher played a lead role in formal analysis, data curation, visualization, and writing–original draft and an equal role in conceptualization, investigation, methodology, and writing–review and editing. Paulo F. Carvalho played a lead role in funding acquisition, an equal role in conceptualization, investigation, methodology, and writing-review and editing.

Funding Open Access funding provided by Carnegie Mellon University. This research was supported in part by National Science Foundation grants #1824257 and #2301130 to PFC, as well as a Google Academic Research Award to PFC.

Data Availability All data and code for this study are openly available at https://osf.io/kruzw. Hypotheses, methods, and the analysis plan for Studies 1 and 2 were pre-registered at https://osf.io/j4rz6 and https://osf.io/28jtr.

Declarations

Ethics Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. All research was approved by the intitutional review board at Carnegie Mellon University.

Consent Informed consent was obtained from all individual participants included in the study.

Conflict of Interest The authors declare no competing interests.

References

Anderson, S., & Rodin, J. (1989). Is bad news always bad?: Cue and feedback effects on intrinsic motivation. Journal of Applied Social Psychology, 19(6), 449–467. https://doi.org/10.1111/j.1559-1816.1989.tb00067.x

Asher, M. W., & Harackiewicz, J. M. (2024). Using choice and utility value to promote interest: Stimulating situational interest in a lesson and fostering the development of interest in statistics. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000921

Asher, M. W., Hecht, C. A., Harackiewicz, J. M., Curtin, J. J., Parrisius, C., & Nagengast, B. (2025). Why elusive expectancy × value interactions may be critical for theory and intervention: A simulated power analysis. *Motivation Science*. https://doi.org/10.1037/mot0000394

Asher, M. W., Sana, F., Koedinger, K. R., & Carvalho, P. F. (2025). Practice with feedback versus lecture: Consequences for learning, efficiency, and motivation. *Journal of Applied Research in Memory and Cognition*. https://doi.org/10.1037/mac0000205

Bandura, A. (1997). *Self-efficacy: The exercise of control*. Freeman.

Bernacki, M. L., & Walkington, C. (2018). The role of situational interest in personalized learning. *Journal of Educational Psychology, 110*(6), 864–881. https://doi.org/10.1037/edu0000250

Bardach, L., Emslander, V., Kasnaci, E., Eitel, A., Lindner, M., & Bailey, D. H. (2025). Research syntheses on AI in education offer limited educational insights [Preprint]. Open Science Framework. https://doi.org/10.31219/osf.io/dx6kt_v1

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth Publishers.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology, 64*(1), 417–444. https://doi.org/10.1146/annurev-psych-113011-143823

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). Taxonomy of educational objectives: The classification of educational goals. In *Handbook 1: Cognitive domain*. Longman New York.

Bohrnstedt, G. W., & Marwell, G. (1978). The reliability of products of two random variables. Sociological Methodology, 9, 254. https://doi.org/10.2307/270812

Bruner, J. S. (1961). The act of discovery. Harvard Educational Review, 31, 21–32.

Butler, A. C., Godbole, N., & Marsh, E. J. (2013). Explanation feedback is better than correct answer feedback for promoting transfer of learning. *Journal of Educational Psychology, 105*(2), 290–298. https://doi.org/10.1037/a0031026

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied, 13*(4), 273–281. https://doi.org/10.1037/1076-898X.13.4.273

Carpenter, S. K., Rahman, S., & Perkins, K. (2018). The effects of prequestions on classroom learning. *Journal of Experimental Psychology: Applied, 24*(1), 34–42. https://doi.org/10.1037/xap0000145

Carvalho, P. F., Asher, M. W., Sana, F., & Koedinger, K. R. (2024). *Skip the reading assignment: Effective and efficient learning with only practice and feedback*. Open Science Framework. https://doi.org/10.31219/osf.io/9y2tb

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. Journal of Applied Psychology, 78(1), 98–104. https://doi.org/10.1037/0021-9010.78.1.98

Costello, T. H., Pennycook, G., & Rand, D. G. (2024). Durably reducing conspiracy beliefs through dialogues with AI. *Science, 385*(6714), eadq1814. https://doi.org/10.1126/science.adq1814

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*(3), 268–294. https://doi.org/10.1037/0096-3445.104.3.268

DeCaro, M. S., & Rittle-Johnson, B. (2012). Exploring mathematics problems prepares children to learn from instruction. Journal of Experimental Child Psychology, 113(4), 552–568. https://doi.org/10.1016/j.jecp.2012.06.009

Durik, A. M., & Harackiewicz, J. M. (2007). Different strokes for different folks: How individual interest moderates the effects of situational factors on task interest. *Journal of Educational Psychology, 99*(3), 597–610. https://doi.org/10.1037/0022-0663.99.3.597

Durik, A. M., Shechter, O. G., Noh, M., Rozek, C. S., & Harackiewicz, J. M. (2015). What if i can't? Success expectancies moderate the effects of utility value information on situational interest and performance. *Motivation and Emotion, 39*(1), 104–118. https://doi.org/10.1007/s11031-014-9419-0

Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology, 61*, 101859. https://doi.org/10.1016/j.cedpsych.2020.101859

Fisher, D., & Frey, N. (2021). *Better learning through structured teaching: A framework for the gradual release of responsibility* (3rd ed.). ASCD.

Glover, J. A. (1989). The "testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology, 81*(3), 392.

Hartwig, M. K., & Dunlosky, J. (2017). Category learning judgments in the classroom: Can students judge how well they know course topics? *Contemporary Educational Psychology, 49*, 80–90. https://doi.org/10.1016/j.cedpsych.2016.12.002

Hecht, C. A., Grande, M. R., & Harackiewicz, J. M. (2021). The role of utility value in promoting interest development. *Motivation Science, 7*(1), 1–20. https://doi.org/10.1037/mot0000182

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. Journal of the Learning Sciences, 21(1), 45–83. https://doi.org/10.1080/10508406.2011.591717

Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology, 19*(4–5), 528–558. https://doi.org/10.1080/09541440601056620

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

Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. Psychometrika, 65(4), 457–474. https://doi.org/10.1007/BF02296338

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The knowledge-learning-instruction framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science, 36*(5), 757–798. https://doi.org/10.1111/j.1551-6709.2012.01245.x

Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), The Cambridge Handbook of Consciousness (1st ed.). Cambridge UniversityPress. https://doi.org/10.1017/CBO9780511816789.012

Koriat, A., & Ackerman, R. (2010). Metacognition and mindreading: Judgments of learning for self and other during self-paced study. *Consciousness and Cognition, 19*(1), 251–264. https://doi.org/10.1016/j.concog.2009.12.010

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory Into Practice, 41*(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2

Linnenbrink-Garcia, L., Durik, A. M., Conley, A. M., Barron, K. E., Tauer, J. M., Karabenick, S. A., & Harackiewicz, J. M. (2010). Measuring situational interest in academic domains. *Educational and Psychological Measurement, 70*(4), 647–671. https://doi.org/10.1177/0013164409355699

Little, J. L., & Bjork, E. L. (2016). Multiple-choice pretesting potentiates learning of related information. *Memory & Cognition, 44*(7), 1085–1101. https://doi.org/10.3758/s13421-016-0621-z

Li, J., King, R. B., Wang, Y., Leung, S. O., & Wang, C. (2023). Students' and schools' expectancy-value beliefs are associated with reading achievement: A cross-cultural study. Learning and Individual Differences, 106, 102344. https://doi.org/10.1016/j.lindif.2023.102344

Loibl, K., & Rummel, N. (2014). The impact of guidance during problem-solving prior to instruction on students' inventions and learning outcomes. Instructional Science, 42(3), 305–326. https://doi.org/10.1007/s11251-013-9282-5

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. Psychological Bulletin, 114(2), 376–390. https://doi.org/10.1037/0033-2909.114.2.376

Mertens, U., Finn, B., & Lindner, M. A. (2022). Effects of computer-based feedback on lower- and higher-order learning outcomes: A network meta-analysis. *Journal of Educational Psychology, 114*(8), 1743–1772. https://doi.org/10.1037/edu0000764

Mook, D. G. (1983). In defense of external invalidity. American Psychologist, 38(4), 379– 387. https://doi.org/10.1037/0003-066X.38.4.379

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In Psychology of Learning and Motivation (Vol. 26, pp. 125–173). Elsevier. https://doi.org/10.1016/S0079-7421(08)60053-5

Pan, S. C., & Carpenter, S. K. (2023). Prequestioning and pretesting effects: A review of empirical research, theoretical perspectives, and implications for educational practice. *Educational Psychology Review, 35*(4), 97. https://doi.org/10.1007/s10648-023-09814-5

Raviv, L., Lupyan, G., & Green, S. C. (2022). How variability shapes learning and generalization. *Trends in Cognitive Sciences, 26*(6), 462–483. https://doi.org/10.1016/j.tics.2022.03.007

Rawson, K. A., Dunlosky, J., & Sciartelli, S. M. (2013). The power of successive relearning: Improving performance on course exams and long-term retention. *Educational Psychology Review, 25*(4), 523–548. https://doi.org/10.1007/s10648-013-9240-4

Renninger, K. A., Bachrach, J. E., & Hidi, S. E. (2019). Triggering and maintaining interest in early phases of interest development. *Learning, Culture and Social Interaction, 23*, 100260. https://doi.org/10.1016/j.lcsi.2018.11.007

Rodriguez-Paz, J. M., Kennedy, M., Salas, E., Wu, A. W., Sexton, J. B., Hunt, E. A., & Pronovost, P. J. (2009). Beyond "see one, do one, teach one": Toward a different training paradigm. *Postgraduate Medical Journal, 85*(1003), 244–249. https://doi.org/10.1136/qshc.2007.023903

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Sabbag, A. G. (2016). *Examining the relationship between statistical literacy and statistical reasoning [doctoral dissertation]*. University of Minnesota.

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. Metacognition and Learning, 4(1), 33–45. https://doi.org/10.1007/s11409-008-9031-3

Stains, M., Harshman, J., Barker, M. K., Chasteen, S. V., Cole, R., DeChenne-Peters, S. E., Eagan, M. K., Esson, J. M., Knight, J. K., Laski, F. A., Levis-Fitzgerald, M., Lee, C. J., Lo, S. M., McDonnell, L. M., McKay, T. A., Michelotti, N., Musgrove, A., Palmer, M. S., Plank, K. M., et al. (2018). Anatomy of STEM teaching in north American universities. *Science, 359*(6383), 1468–1470. https://doi.org/10.1126/science.aap8892

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

Walkington, C. A. (2013). Using adaptive learning technologies to personalize instruction to student interests: The impact of relevant contexts on performance and learning outcomes. *Journal of Educational Psychology, 105*(4), 932–945. https://doi.org/10.1037/a0031882

Wang, M.-T. (2012). Educational and career interests in math: A longitudinal examination of the links between classroom environment, motivational beliefs, and interests. *Developmental Psychology, 48*(6), 1643–1657. https://doi.org/10.1037/a0027247

Zepeda, C. D., Martin, R. S., & Butler, A. C. (2020). Motivational strategies to engage learners in desirable difficulties. *Journal of Applied Research in Memory and Cognition, 9*(4), 468–474. https://doi.org/10.1016/j.jarmac.2020.08.007

Authors and Affiliations

Michael W. Asher[1] · Paulo F. Carvalho[1]

✉ Michael W. Asher
masher@andrew.cmu.edu

[1] Human-Computer Interaction Institute, Carnegie Mellon University, Newell-Simon Hall, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA