# Journal of Applied Research in Memory and Cognition

Practice With Feedback Versus Lecture: Consequences for Learning, Efficiency, and Motivation

Michael W. Asher, Faria Sana, Kenneth R. Koedinger, and Paulo F. Carvalho

Online First Publication, January 27, 2025. https://dx.doi.org/10.1037/mac0000205

Journal of Applied Research in Memory and Cognition

ISSN: 2211-3681

https://doi.org/10.1037/mac0000205

EMPIRICAL ARTICLE

Practice With Feedback Versus Lecture: Consequences for Learning, Efficiency, and Motivation

Michael W. Asher[1], Faria Sana[2], Kenneth R. Koedinger[1], and Paulo F. Carvalho[1]
[1] Human-Computer Interaction Institute, Carnegie Mellon University, United States
[2] Centre for Social Sciences, Athabasca University, Canada

Many college students drop science, technology, engineering, and math (STEM) majors after struggling in gateway courses, in part because these courses place large demands on students' time. In three online experiments with two different lessons (measures of central tendency and multiple regression), we identified a promising approach to increase the efficiency of STEM instruction. When we removed lectures and taught participants exclusively with practice and feedback, they learned at least 15% faster. However, our research also showed that this instructional strategy has the potential to undermine interest in course content for less confident students, who may be discouraged when challenged to solve problems without upfront instruction and learn from their mistakes. If researchers and educators can develop engaging and efficacy-building activities that replace lectures, STEM courses could become better learning environments.

General Audience Summary

Many college students drop out of science, technology, engineering, and math (STEM) majors after struggling in gateway courses, in part because these courses place large demands on students' time. To address this issue, which affects students with outside-of-school responsibilities (e.g., jobs, families) in particular, instructors must strive to make their courses efficient learning opportunities. Decades of research show that students learn most effectively through active practice, rather than by listening, yet instructors often invest substantial class time lecturing about course content before students practice with activities, problem sets, and study sessions. Must students learn from direct instruction before they can try things themselves and learn by doing? In three online experiments, we tested if participants could effectively and efficiently learn about statistics through practice and feedback, and we also tracked the motivational consequences of this form of instruction. Results showed that when we removed lectures and taught participants exclusively with practice and feedback, they learned at least 15% faster. However, we also found that this instructional strategy decreased interest in statistics for less confident students, who may have been discouraged when challenged to solve problems without upfront instruction. In sum, this research provides initial evidence that students may be able to learn more efficiently when lectures are replaced with practice opportunities and feedback, but careful work is needed to (a) design engaging and motivating practice-based instruction and (b) evaluate this instructional approach in real STEM courses. If researchers and educators can develop engaging and efficacy-building activities that replace lectures, STEM courses may become better learning environments.

Keywords: active learning; practice testing; science, technology, engineering, and math education

Supplemental materials: https://doi.org/10.1037/mac0000205.supp

Sean Kang served as action editor.
Michael W. Asher https://orcid.org/0000-0002-1006-8813
Faria Sana https://orcid.org/0000-0002-2202-7592
Kenneth R. Koedinger https://orcid.org/0000-0002-5850-4768
Paulo F. Carvalho https://orcid.org/0000-0002-0449-3733
All data and code for this study are openly available at https://osf.io/3ghwv. Study 1 was not preregistered. Hypotheses, methods, and the analysis plan for Studies 2 and 3 were preregistered at https://osf.io/hkaxu and https://osf.io/mnx8z. For both studies, there were minor deviations from the preregistrations (for details, see the Supplemental Material section titled "Measures and Deviations from the Preregistration" for each study). The findings of Studies 1 and 2 are also presented in the conference proceedings for the 2024 meeting of the Cognitive Science Society. All other authors declare no conflicts of interest.

This research was supported in part by National Science Foundation Grants 1824257 and 2301130 to Paulo F. Carvalho. The funding source had no other role than financial support.

continued

Every year, many aspiring science, technology, engineering, and math (STEM) majors switch their fields of study or even drop out of college after struggling in introductory math and science courses (Chen, 2013; Rosenzweig et al., 2021). One cause of this problem is that STEM courses place large demands on students' time, a cost that affects students with outside-of-school responsibilities (e.g., jobs, families) in particular. To address this issue, instructors must strive to make their courses efficient learning opportunities. Fortunately, STEM courses may have substantial room to improve how they allocate students' time. Instructors often invest substantial class time lecturing about course content (Laursen, 2019; Stains et al., 2018), which students then practice with activities, problem sets, and study sessions (Freeman et al., 2014). Do students need both lecture and practice to succeed, or might eliminating one method of instruction (and focusing more on the other) yield more efficient learning?

The importance of practice is well-established. Research on topics like the testing effect, prequestioning, active learning, and deliberate practice shows that students master skills and acquire knowledge most effectively when they actively work with the relevant information, testing their understanding, receiving feedback that allows them to correct mistakes, and practicing correct responses (Carpenter et al., 2022; Ericsson et al., 1993; Freeman et al., 2014; Koedinger et al., 2015; Macnamara et al., 2014; Pan & Carpenter, 2023; Roediger & Karpicke, 2006). Even when students spend large amounts of time on forms of explicit instruction like lectures and readings, they typically still require practice opportunities to master academic skills (Koedinger et al., 2023).

Surprisingly, the necessity of lectures is less clear. Although the importance of guided instruction is well-established (Kirschner et al., 2006), there is little experimental evidence about whether a lecture is needed if participants are already learning from carefully structured practice and feedback. There is good reason to expect that a combination of lecture, practice, and feedback would be superior to practice and feedback alone. Students may struggle to learn from problems with many unlearned elements that overwhelm working memory (see cognitive load theory; e.g., Kirschner, 2002). In this case, an upfront lecture can orient students toward relevant information, helping them organize and effectively store it in memory (Sweller, 2004). Subsequently, when students attempt practice problems, they may be more likely to reinforce correct responses. In addition, repetition usually has a positive effect on learning and retention, even with passive instructional approaches (Rothkopf, 1968).

However, it is also possible that lectures are redundant and inefficient if students already have access to practice opportunities and feedback. The Knowledge-Learning-Instruction (KLI) framework (Koedinger et al., 2012) models the theoretical relationship between instruction, learning, knowledge, and assessment. In this framework, instructional events are intended to bring about unobservable learning events which update students' knowledge. Knowledge is inferred with assessment events (e.g., homework assignments, essays, tests, discussions). When students passively learn about a fact or skill via lecture, they complete an instructional event. However, when students attempt to solve a problem and receive feedback about that same information, they complete both an instructional event (the feedback) and an assessment event (the question).

Critically, there is reason to believe that this combination of instruction and assessment leads to better learning. For instance, the assessment aspect of the feedback has metacognitive benefits, providing learners with information about their own knowledge states, highlighting important information that they know and that they have not yet successfully learned, and decreasing overconfidence (R. A. Bjork et al., 2013; Pan & Carpenter, 2023). When the learner subsequently attends to the instructional component of the feedback (the correct response), they can better focus on relevant information and process it more deeply, updating their knowledge (Koedinger et al., 2012). Therefore, it is possible that direct instruction is not the only tool that can help students manage their cognitive load. Well-scaffolded practice problems and feedback should also help students focus on relevant and manageable chunks of information, promoting effective learning.

Evidence for this process can be found in studies of "prequestions" in which instructional events like written passages or lectures are preceded by questions about their contents. For example, Carpenter and Toftness (2017) manipulated whether participants completed a set of short-answer practice questions, or not, before they watched a recorded history lecture. No immediate feedback was provided; the researchers reasoned that participants would more attentively process the subsequent lecture to learn about the prequestioned information (and the correctness of their responses). As predicted, on an end-of-session test, participants who completed prequestions performed a full standard deviation better than those in the control group. Practice with immediate feedback can be seen as a form of prequestions where the feedback replaces the text or lecture. Thus, completing prequestions followed by feedback should help students attend to and learn from the feedback, even without a subsequent lecture (Pan & Carpenter, 2023).

Analyses of student behavior in online courses also provide evidence that lecture may not be needed when students can learn via practice and feedback. In these courses, researchers find that when students choose to invest time completing activities and receiving feedback, they learn much more than they do by reading and watching videos (Carvalho et al., 2022; Koedinger et al., 2015). If students can learn effectively from practice and feedback, instructors may be able to make their courses more efficient learning opportunities by focusing on this type of instruction and removing redundant lectures. In the present research, we conducted a series of three online experiments to collect initial evidence about this hypothesis.

Michael W. Asher played a lead role in formal analysis, visualization, and writing–original draft and an equal role in conceptualization, data curation, investigation, methodology, and writing–review and editing. Faria Sana played a supporting role in investigation and methodology and an equal role in conceptualization. Kenneth R. Koedinger played an equal role in conceptualization. Paulo F. Carvalho played a lead role in funding acquisition and validation and an equal role in conceptualization, data curation, investigation, methodology, and writing–review and editing.

The data are available at https://osf.io/3ghwv/?view_only=c62a272345a54f51a084cb534d0ad7a5.

The experimental materials are available at https://osf.io/3ghwv/?view_only=c62a272345a54f51a084cb534d0ad7a5.

Correspondence concerning this article should be addressed to Michael W. Asher, Human-Computer Interaction Institute, Carnegie Mellon University, Newell-Simon Hall, 5000 Forbes Avenue, Pittsburgh, PA 15213, United States. Email: masher@andrew.cmu.edu

Study 1: Can Students Learn Statistics From Practice, Without Lecture?

We conducted Study 1 to investigate whether participants can learn from practice-based instruction (practice with feedback only) and compare this approach to standard STEM instruction: a combination of lecture and practice with feedback. In this study, participants learned about a statistics topic and then took a test. Participants were randomly assigned to one of four conditions: no instruction, practice with feedback only, lecture-only, or combined instruction.

We predicted that practice with feedback would be an effective instructional event, such that participants in the practice condition would learn more than those in both the no instruction and lecture-only conditions. We also reasoned that two instructional events (practice with feedback, plus lecture) would be better than one (practice with feedback only), and therefore predicted that participants in the combined condition would learn more than those in the practice-only condition. However, we predicted that practice-based instruction would result in much more efficient learning.

In addition, we were concerned that even though practice-based instruction would promote efficient learning, participants would not appreciate the benefits of this instructional strategy. As students follow along with a lecture without having their understanding challenged, they are likely to experience a sense of fluency, comprehension, and therefore confidence (E. L. Bjork & Bjork, 2011). Conversely, when students test their understanding with practice questions, they may struggle and feel as though they learned less (Kirk-Johnson et al., 2019). To investigate this concern, we assessed participants' subjective judgments of learning as a secondary outcome.

Method

For all studies, we report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study.

Participants

A total of 132 participants were recruited through Prolific who consented to participate and completed the study. On Prolific, participants were screened to be ≥18 years old and living in the United States. Ninety-seven (79%) self-reported their ethnicity as White, 14 (11%) as Black, 10 (8%) as Asian, four (3%) as multiracial, and two (2%) as belonging to another group. Two participants (2%) chose not to report their ethnicity. Seventy-nine participants (60%) identified as women and 50 (28%) as men. Three participants chose not to report their gender (2%). The average age of participants was 41.6 years. The study lasted approximately 30 min, and participants were paid $4.80 for their participation. The sample size was determined to detect effects $f > .3$ with 80% power, which we determined based on a prior experiment that we report in our Supplemental Material.

Procedure

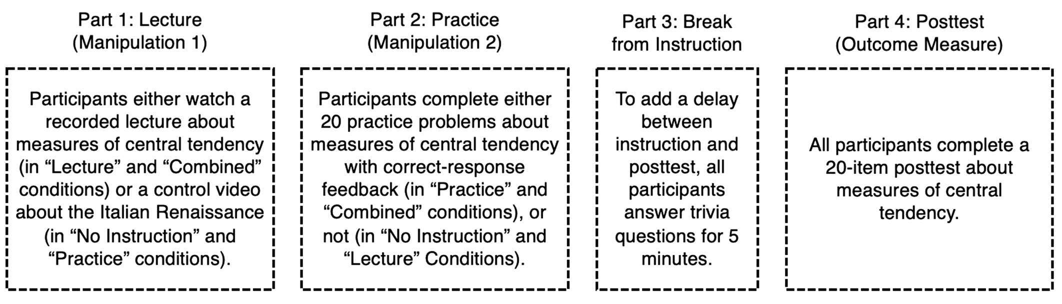
The study contained four sections: lecture, practice, a 5-min break, and then a test (see Figure 1). During the lecture section, participants who were randomized to the "lecture" and "combined" conditions watched a recorded lecture about measures of central tendency. All other participants watched a control lecture, which had a similar style and took a similar amount of time but covered an unrelated topic (the Italian Renaissance). Next, participants who were randomly assigned to the "practice" and "combined" conditions completed a 20-question practice test, receiving correct-response feedback after each question (i.e., the word "Correct" or "Incorrect," followed by the correct answer). All other participants skipped this practice test. Third, all participants completed a series of trivia questions for 5 min. Finally, all participants proceeded to a survey and then a posttest about measures of central tendency.

Materials

For the recorded lecture, we used an 11.4-min educational video about measures of central tendency that was developed by the YouTube channel CrashCourse. In the video, the instructor spent 63% of the time defining terms and explaining concepts, 25% working through sample problems, 8% explaining the relevance of the content, and 5% transitioning between topics. The control video was a 14.6-min CrashCourse lecture about the Italian Renaissance.

To create practice and assessment materials, we developed 20 "knowledge" questions that tested facts that were covered in the lecture (e.g., definitions), and 20 "application" questions that required participants to apply the information. These questions were developed to correspond to 16 primary learning objectives in the lecture, identified by the research team, with at least one question per objective. Learning objectives and corresponding questions are provided in the Supplemental Material.

Figure 1
Timeline of Study 1



Because individuals best learn to generalize when they are exposed to varied input (Raviv et al., 2022), we expected that the variety of combined instruction may be most important for application questions. We then randomly split the knowledge and application items into two problem sets, Versions A and B. Participants were randomly assigned to a version, which they received for practice (if applicable) and for the posttest. We originally intended to counterbalance the order of the tests, but due to a coding error, participants in practice conditions were given the same test at both times. All questions were four-item multiple choice, and the order of questions was randomized for each participant. The trivia questions given during the 5-min break were developed and normed by Tauber et al. (2013).

In addition, we tracked instructional time for each participant (i.e., the time they spent on the lecture and/or practice sections; $M = 9.2$ min, $SD = 7.2$ min). We inserted a single item after instruction to assess participants' self-reported judgments of their learning ("The instruction I just received prepared me well to answer questions about measures of central tendency"), adapted from Koriat and Ackerman (2010), $M = 55.1$, $SD = 40.2$. Participants responded to this item using a slider that ran from 0 (*strongly disagree*) to 100 (*strongly agree*). Following a 5-min break and posttest ($M = 70.0\%$, $SD = 20.4\%$), participants were asked which form of instruction would have been best: lecture, practice with feedback, or a combination of lecture, practice, and feedback.

To estimate the efficiency of instruction for each participant, three pieces of information are needed: end-of-session knowledge, initial knowledge, and instruction time. We directly measured end-of-session knowledge with the posttest, and we timed how long the lecture and/or practice took for each participant, but we chose not to include a pretest (to measure initial knowledge) because we were concerned that a pretest, even without feedback, would have metacognitive benefits and thereby increase learning in the lecture-only condition (see Carpenter & Toftness, 2017; Sana et al., 2020). Instead, we estimated initial knowledge for participants in the practice-only condition using their performance at the beginning of the practice session, $M = 54.3\%$, $SD = 16.2\%$.

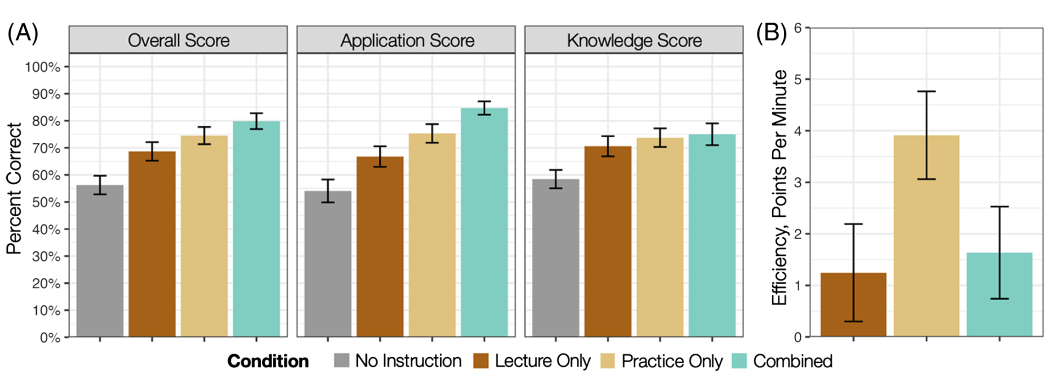
Because initial knowledge could not be estimated for participants in the lecture-only and combined conditions (which lacked a practice session that came before the lecture), we used the initial knowledge estimates from the practice-only condition and a resampling procedure to generate 1,000 imputed data sets, each with different plausible estimates for students' initial knowledge in all conditions. We then calculated efficiency scores for participants in each imputed data set and analyzed the data sets separately, pooling regression estimates to yield an unbiased estimate of efficiency while reflecting the uncertainty in participants' true initial knowledge (Rubin, 1987). This process is detailed in the Supplemental Material.

Results

All analyses were conducted using R Version 4.3.1 (R Core Team, 2023) with the "lmer" package (Bates et al., 2015). To test our hypotheses, we calculated each participant's scores on knowledge and application questions on the posttest, and then fit a series of three linear mixed-effects models, each one regressing knowledge and application scores on a set of dummy-coded contrasts to ultimately test each possible pairwise comparison between the four conditions. In each model, we also included an *Application versus Knowledge* contrast, which indicated whether each score was for application (.5) or knowledge (−.5) questions, a *Version* contrast to control for whether participants took Version "A" (.5) or "B" (−.5) of the final test, the two-way interactions between the *Application versus Knowledge* contrast and the dummy-coded condition contrasts, a by-participant random intercept, and a by-participant random slope for the *Application versus Knowledge* contrast. Degrees of freedom and p values were calculated using a Kenward–Roger F test and adjusted for multiple comparisons to keep the false discovery rate ≤5% (Benjamini & Hochberg, 1995; Kenward & Roger, 1997). Because judgments of learning were a between-subjects variable, we analyzed them with the same approach used for performance but with multiple regression. To analyze efficiency in the three conditions that contained instruction, we fit a regression model with dummy-coded contrasts that compared the combined and lecture-only conditions to practice-only, controlling for version. In the Supplemental Material, for Studies 1–3, we report correlations between all measures, descriptive statistics, and detailed output from all regression models.

Figure 2 displays average performance and efficiency by condition.

Figure 2
Posttest Scores and Efficiency of Learning, Study 1



Note. A displays means by condition; error bars show ±1 standard error of each mean. B displays predicted values from the model we fit to estimate efficiency in each condition; error bars represent ±1 standard error of each estimate. See the online article for the color version of this figure.

Performance

The left panel of Figure 2A shows average overall performance by condition. Compared to those in the no instruction condition, participants in the practice-only conditions performed on average 19 points better on the posttest, $b = 0.19$, $d = 1.03$, $F(1, 127) = 16.95$, $p < .001$, demonstrating substantial learning without the need for lecture. There was no significant difference between the performance of participants in the practice-only and lecture-only conditions, although this effect was in the predicted direction: Participants in the practice-only condition performed on average 5 points better on the posttest, $b = 0.05$, $d = 0.29$, $F(1, 127) = 1.17$, $p = .266$. There was also no significant difference in performance between the combined and practice-only conditions, although participants in the combined condition performed on average 8 points better, $b = 0.08$, $d = 0.43$, $F(1, 127) = 2.92$, $p = .135$.

The second and third panels of Figure 2A display average performance on application and knowledge items. Two significant interactions with the *Application versus Knowledge* contrast emerged, suggesting that combined instruction was particularly good at promoting performance on application (vs. knowledge) questions, relative to no instruction, $b = 0.14$, $F(1, 128) = 10.13$, $p = .007$, and lecture-based instruction, $b = 0.14$, $F(1, 128) = 9.64$, $p = .007$. No other interactions with the *Application* contrast were significant, $p \geq .117$.

Efficiency

Figure 2B shows the estimated average efficiency of learning in each condition that contained instruction. As predicted, participants learned at an estimated rate of 3.91 points per minute in the practice-only condition, 2.4× more efficiently than those in the combined condition (1.63 points per minute), $b = -2.28$, $t(87.25) = -3.58$, $p = .001$, and more than 3× more efficiently than those in the lecture-only condition (1.24 points per minute), $b = -2.67$, $t(80.99) = -4.11$, $p < .001$.

Judgments of Learning and Instruction Preferences

Figure 3A displays average judgments of learning by condition. Although practice improved posttest performance by 19 points relative to no instruction, it had no effect on participants' subjective judgments of learning, $b = -4.48$, $d = -0.15$, $t(117) = -0.60$, $p = .662$. Similarly, participants in the combined condition reported similar judgments of learning to those in the lecture-only condition, $b = -2.38$, $d = -0.08$, $t(117) = -0.32$, $p = .747$, despite performing 13 points better on the posttest. In addition, when participants were asked at the end of the study which form of instruction would be best for preparing them for a test, combined instruction was chosen by 70% of participants, followed by lecture-only (17%). Although it was three times as efficient as lecture and resulted in higher test scores, practice-only was the least popular option, selected as the best method of preparation by only 12% of participants (Figure 3B).

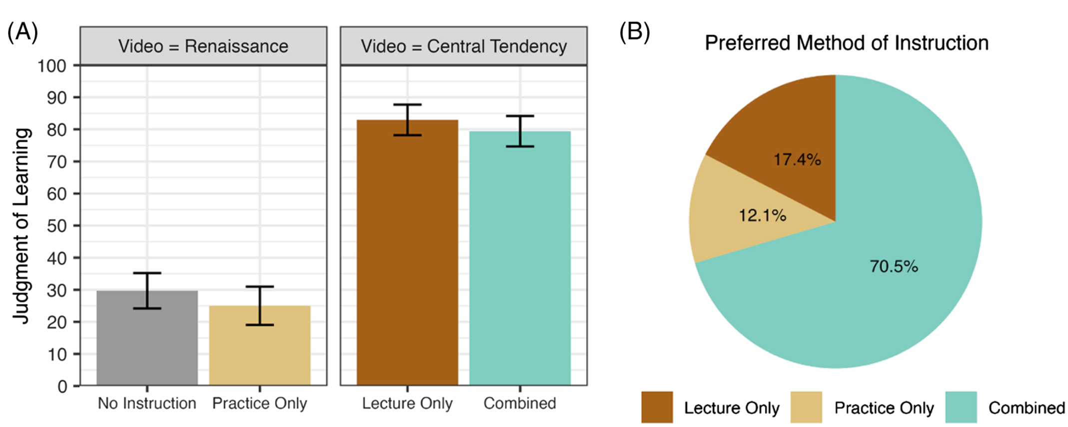
Discussion

Study 1 showed that participants could learn effectively from practice and feedback, without the need for lecture. Participants in the practice-only condition performed more than a full standard deviation better than those who received no instruction. Although we found some evidence that combined instruction was most effective at promoting performance on application questions, possibly due to the additional variety of information it contained (Raviv et al., 2022), this benefit was offset by a large loss in efficiency. Participants in the practice-only condition learned more than twice as quickly as those in the combined condition, and more than 3× faster than those in the lecture-only condition. However, two major limitations of Study 1 raise questions about the generalizability of our findings to authentic educational contexts.

The first limitation is the narrowed focus of our practice-based instruction. Although we wrote 40 practice questions that covered 16 learning objectives identified in the central tendency lecture, when we randomly divided these questions into two problem sets, neither set covered every objective: Set A covered only 10 and Set B covered 11. Second, participants in the practice conditions were given the same problem set for both the practice and test. Together, these limitations mean that during instruction, participants in the practice-only condition were exposed exclusively the relevant learning objectives that would appear on the test. If the practice were matched to the lecture, rather than the test, it is possible that it would

Figure 3

Judgments of Learning and Instruction Preferences, Study 1



Note. (A) Bars represent average judgments of learning for each condition, and error bars show ±1 standard error of each mean. (B) The proportion of participants who reported a preference for each method of instruction. See the online article for the color version of this figure.

be less efficient and effective. In addition, because of the match between the practice and test, participants who received practice-based instruction could solve application questions on the test by memorizing answers, rather than generalizing from knowledge and skills they had learned. It is critical to address these limitations before drawing conclusions about benefits of practice-based instruction.

In Study 1, we also observed that although practice-based instruction was more effective than lecture, participants judged that it did not prepare them for the test, and a large majority of participants indicated a preference for combined instruction. This raises the concern that although practice-based instruction can be good for learning, it may undermine student motivation. Study 2 investigates this possibility and addresses the limitations of Study 1.

Study 2: Can Practice With Feedback Promote Generalization and Motivation?

In Study 2, we set out to replicate and extend the findings of Study 1 in a new context, with new materials and actual college students. In addition to focusing on learning and efficiency of instruction as outcomes, we examined how practice-based instruction affects motivation in a context that more closely resembles a college-level STEM course. We designed an experimental paradigm in which participants learned the basics of multiple regression, as they would in an introductory statistics course, either with a lecture or with practice problems and feedback. To increase the potential effectiveness and efficiency of the lecture, we built it around two worked examples in which the instructor used regression to answer research questions (Atkinson et al., 2000). We designed the practice and feedback so participants would work through the same worked examples in the same order as the lecture, receiving elaborated feedback about each correct answer and why it was correct. As such, the practice and feedback covered the same information as the lecture not the test, addressing a major limitation of Study 1. After instruction, all participants took an application-heavy test of items that were new for all participants, testing their ability to generalize.

Based on the results of Study 1, we predicted that participants would learn more from practice than lecture in less time. However, because every major theory of academic motivation involves students' beliefs about their own competence (Bandura, 1986; Eccles & Wigfield, 2020; Ryan & Deci, 2000), and because learning via practice involves struggle and negative feedback, we predicted that a practice-based instructional approach might undermine participants' judgments of learning, confidence, and interest in statistics. This study was preregistered at https://osf.io/hkaxu.

Method

Participants

Undergraduate participants were recruited to participate in this study from an introductory psychology course at a large midwestern university. A total of 338 students consented to participate and completed the study. Of these participants, 225 (67%) self-reported their ethnicity as White, 75 (22%) as Asian, 32 (9%) as Hispanic, 14 (4%) as Black, four (1%) as Middle Eastern, three (1%) as Indigenous, and one (<1%) as belonging to another group. Two hundred twelve participants (63%) identified as women and 123 (36%) as men. Two participants identified as nonbinary (<1%) and one did not report their gender (<1%). The average age of participants was 18.5 years. Participants completed the session online for course credit. Data collection lasted for the duration of the Fall semester.

Procedure

Study 2 had a two-cell design with lecture- and practice-only conditions. Before the learning session began, all participants completed measures of their baseline confidence and interest in statistics. Next, participants either watched a 13.8-min lecture that we recorded or completed a series of 16 practice problems. After each problem, participants received feedback. The lecture was built around a series of two worked examples, which occupied 51% of the lecture. In addition, the instructor spent 32% of the lecture on definitions and explanations, 12% on transitions between topics, and 5% on an end-of-lecture recap. Unlike the prior study, there was no control video in the practice-only condition to facilitate an unconfounded test of how the two manipulations affected students' interest. With the practice problems, all participants worked through the same information in the same order.

Finally, participants in both conditions completed the same 21-question posttest, which was designed with ecological validity in mind to closely resemble a college-level statistics test. Specifically, the test consisted of four parts, which were presented in the same order to all participants. Part 1 included four multiple-choice knowledge questions that tested memory of facts that were presented during instruction. Parts 2–4 were each built around a different equation or graph (e.g., a scatterplot of real data about the relationship between gas prices and traffic fatalities) and contained 17 short-answer application questions. All instructional materials are shared at https://osf.io/3ghwv (Asher & Carvalho, 2024).

Measures

Baseline interest in statistics was measured with three items (e.g., "How interesting do you find statistics?" $\alpha = .93$), as was baseline confidence in math (e.g., "How good are you at math?" $\alpha = .93$). On the outcome questionnaire, interest in statistics was measured with 12 items (e.g., "How interesting do you find linear regression?"; "How much would you enjoy learning more about statistics in the future," $\alpha = .96$), and we measured confidence in regression with three items (e.g., "How well do you think you would do in a regression course?" $\alpha = .93$). These scales were adapted from Linnenbrink-Garcia et al. (2010), Durik et al. (2015), and Hecht et al. (2021). In addition, participants received an overall score on the posttest (M = 54%, SD = 21%), and we calculated subscores on knowledge (M = 75%, SD = 33%) and application (M = 49%, SD = 25%) questions for an exploratory analysis of performance on these different types of questions. Because Study 2 lacked counterbalanced practice and posttests, measures of learning per minute could not be calculated. Instead, we use the amount of time students spent on instruction as a measure of efficiency. We assessed participants' judgments of their learning with the same item used in Study 1.

Results

We regressed each outcome on a *Practice versus Lecture* contrast (Practice = .5, Lecture = −.5), our baseline measures of interest and confidence (both standardized), and interactions between the *Practice* contrast and the two baseline measures. We predicted that practice could undermine confidence and interest for less confident students (who might be threatened when asked to practice without any upfront instruction), and we wanted to examine if less interested students would prefer lecture or practice. In an exploratory (i.e., nonpreregistered) analysis, we fit a linear mixed-effects model to test if *Practice versus Lecture* effects varied on the knowledge and application questions.

Performance

Contrary to our prediction, participants in both conditions performed similarly on the test. Although participants who learned from practice once again did better on the test than those who learned from lecture, this difference was small (only 3 percentage points, $d = 0.12$) and not statistically significant, $b = 0.03$, $t(332) = 1.22$, $p = .225$, Figure 4A. There were no significant interactions with baseline interest or confidence, $p \geq .071$. In our exploratory analysis of application versus knowledge subscores, we found a significant *Practice Versus Lecture* × *Application Versus Knowledge* interaction, $b = 0.07$, $t(336) = 4.83$, $p = .029$, suggesting that practice-based instruction was more effective at preparing students for application questions than knowledge questions, Figure 4B. Notably, this pattern was also observed when comparing the practice and lecture conditions in Study 1, although it was nonsignificant and slightly smaller, $b = 0.04$.

Efficiency

Again, practice was a much more efficient form of instruction than lecture. Whereas participants took 15.1 min on average to watch the lecture, they averaged only 6.7 min to complete the practice problems, $b = -7.94$, $t(332) = -31.30$, $p < .001$, Figure 4C. Assuming that participants in both conditions learned a similar amount, this translates to more than a 2.25× increase in the efficiency of instruction. There were no significant interactions with baseline interest or confidence, $p \geq .489$.

Judgment of Learning

There was no significant difference in the extent to which participants in the two conditions reported that their form of instruction prepared them for the test, although the effect on this outcome was in the predicted direction: Judgments of learning were 0.14 *SD* lower in the practice condition than in the lecture condition, $b = -0.18$, $d = -0.14$, $t(332) = -1.43$, $p = .154$. There were no significant interactions with baseline interest or confidence, $p \geq .236$.

Confidence in Regression Ability

Participants in both conditions reported similar levels of confidence in their regression ability, with students in the practice condition reporting slightly higher levels, $b = 0.11$, $d = 0.12$, $t(332) = 1.23$, $p = .220$. This result ran contrary to our concern that the challenges students faced while practicing might undermine students' ability beliefs. There were no significant interactions with baseline interest or confidence, $p \geq .790$.

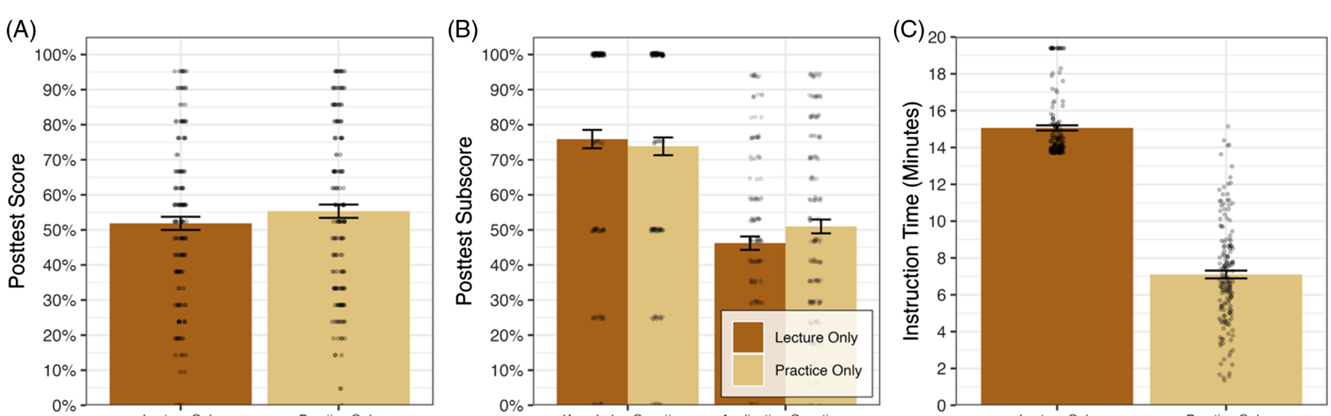
Interest in Statistics

In the two conditions, after instruction, participants reported similar levels of interest in statistics on average, $b = -0.02$, $d = -0.02$, $t(332) = -0.18$, $p = .860$. However, a significant *Practice Versus Lecture* × Baseline Interest interaction suggested that whereas the lecture condition led students at all levels of confidence to report a moderate level of interest in statistics, the practice condition promoted interest for participants with higher levels of confidence and undermined interest for those who were less confident, $b = 0.23$, $t(332) = 2.13$, $p = .034$, Figure 5. There was no significant interaction with baseline interest in statistics, $p = .678$.

Discussion

After participants learned either from a lecture or a closely matched set of practice problems with feedback, they performed similarly well on a posttest. There was no evidence that instruction with practice and

Figure 4
Posttest Scores and Instructional Time, Study 2



Note. (A) Bars represent average posttest scores. (B) Average posttest knowledge and application subscores. (C) Average instruction time for participants in each condition. Error bars show ±1 standard error of each mean. Each dot represents an individual participant, jittered on the *x*-axis to show dispersion. See the online article for the color version of this figure.

feedback limited participants' ability to generalize on the final test; the benefits of practice-based instruction (vs. lecture) were actually largest on questions where participants applied what they had learned to solve novel problems. Because Study 2 lacked a "combined" (lecture plus practice with feedback) condition, we were unable to test whether combined instruction might be even more effective at promoting generalization, as we saw in Study 1. In addition, we again demonstrated that practice with feedback was much more efficient than lecture; participants achieved comparable performance in less than half the time. Thus, Study 2 addresses the limitations of Study 1 and suggests (a) that practice with feedback can be efficient even when it covers all the material in a lecture and (b) that it can effectively teach students to generalize as well as memorize.

However, our concern that practice-based instruction might undermine interest was also supported: Although this type of instruction was better than lecture for confident students (who may have appreciated the challenge of the practice problems and viewed it as appropriate), it undermined interest for less confident students who may have been overwhelmed by the challenge and incorrect-response feedback. In a third and final study, we returned to the central tendency paradigm to replicate our findings about generalization, efficiency, and motivation with new practice conditions and materials.

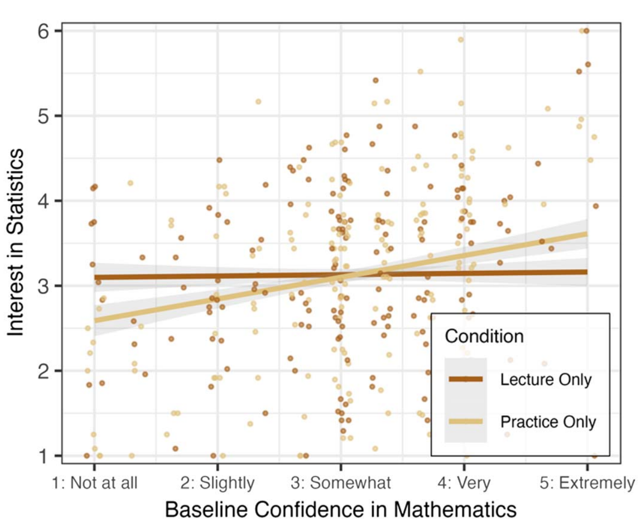
Study 3: Compared to Combined Instruction, Can Matched Practice and Feedback Efficiently Promote Generalization and Motivation?

In Studies 1 and 2, participants learned just as much from practice with feedback as they did from lecture, in 2–3× less time. However, in Study 1, we also found that combined instruction (i.e., lecture plus practice with feedback) led to the best overall performance on application questions. This is possibly because it provided participants with additional, varied exposures to each learning objective. What if an instructor matched combined instruction with additional practice and feedback? That is, what if students learned from two sets of practice problems: a problem set matched to the lecture, followed by the practice problems given in combined instruction? Would this type of matched practice be just as effective (or more effective) at promoting generalization? And might it remain more efficient?

And what about motivation? In both combined and practice-based instruction, students practice and receive (often critical) feedback. However, students may respond more positively when a lecture helps prepare them for the practice, rather than learning exclusively from their mistakes. In addition, participants who skip the lecture will miss features that might be motivationally beneficial. For instance, in the recorded central tendency lecture, the instructor spends 51 s discussing the relevance of the material, a teaching practice that can promote interest (Asher & Harackiewicz, 2024; Canning & Harackiewicz, 2015; Gaspard et al., 2015).We tested these questions in Study 3, which was preregistered at https://osf.io/mnx8z.

Figure 5

Interest in Statistics by Condition and Baseline Confidence in Mathematics



Note. Predicted values are graphed. Error envelopes represent ±1 standard error. See the online article for the color version of this figure.

Method

Participants

On Prolific, participants were screened to be ≥18 years old and living in the United States. A total of 400 participants were recruited through Prolific who consented to participate and completed the study, based on the power analysis included in our preregistration for this study. Of these participants, 236 (59%) self-reported their ethnicity as White, 73 (18%) as Black, 43 (11%) as Asian, 25 (6%) as multiracial, and 20 (5%) as belonging to another group. Three participants (<1%) chose not to report their ethnicity. One hundred fifty-three participants (38%) identified as women and 247 (62%) as men. The average age of participants was 40.4 years. The study lasted approximately 30 min, and participants were paid $6.00 for their participation.

Procedure and Materials

Participants were randomly assigned to one of three conditions: no instruction, combined instruction (a combination of lecture and practice with feedback), or a new, matched version of practice-based instruction (consisting of two practice sessions with feedback). The instructional materials for the no instruction and combined instruction conditions were identical to those used in Study 1. However, for the matched practice condition, we developed a new set of 29 practice problems to closely match the central tendency lecture, covering all the same learning objectives and providing the same examples (in practice question form) to demonstrate skills and illustrate concepts (see the Supplemental Material for details). After completing this problem set, participants in the matched practice condition completed the same practice problems as participants in the combined instruction condition. At the end of the study, all participants completed a posttest consisting of 10 application questions. We also counterbalanced the order of the problem sets given during the second practice section and the posttest to ensure that no participant answered the same question twice.

In addition, we administered brief questionnaires at the beginning of the study, after instruction, and before the test to assess participants' baseline interest and confidence (Questionnaire 1), their interest in the instructional session and statistics (Questionnaire 2), and their confidence and judgments of learning (Questionnaire 3). Questionnaire items were identical to those used in Study 2, except (a) they asked about statistics rather than linear regression wherever applicable and (b) we added three additional two-item measures to better understand the motivation-related consequences of learning from practice and feedback versus combined instruction. We expected that completing practice questions would be a more engaging task than watching a lecture, so we measured self-reported distraction during the learning session (e.g., "I got distracted as I learned during this session." $\alpha = .89$). We added a measure of participants' beliefs about the usefulness of the topic (e.g., "How useful is it to know about measures of central tendency?" $\alpha = .91$) to monitor a potential negative consequence of shifting away from the lecture, which discusses the utility of the material. We also measured participants' confidence about the upcoming test (e.g., "How well do you think you will do on the test?" $\alpha = .89$), which we reasoned might be more sensitive to our manipulations than the more general measure of confidence that we gave in Study 2. Because (like Study 2) Study 3 lacked counterbalanced pre- and posttests, we used the time participants spent on instructional materials (videos and practice questions) as our measure of efficiency. The procedure for Study 3 is summarized in Figure 6.

Results

In all models, we tested the effects of condition with two orthogonal contrasts: An *Instruction versus No Instruction* contrast compared the average of both types of instruction to the control condition and a *Practice versus Combined Instruction* contrast compared the two types of instruction head-to-head. As we did in Study 2, we interacted the condition contrasts with participants' baseline confidence and baseline interest in statistics (standardized). We also controlled for the version of the posttest that participants took. We predicted that practice could undermine confidence and interest for less confident students (who might be threatened when asked to practice without any upfront instruction), and we wanted to explore if less interested students would prefer lecture or practice.

Performance

On average, participants who received either type of instruction outperformed those in the control conditions on the test by 7 points, $b = 0.07$, $d = 0.32$, $t(390) = 2.43$, $p = .015$, Figure 7A. In addition, there was no significant difference in performance between the two types of instruction; participants who completed matched practice performed only 1 point better on the test than those who completed combined instruction, $b = 0.01$, $d = 0.04$, $t(390) = 0.39$, $p = .698$. There were no significant interactions with baseline interest or confidence, $p \geq .676$.

Efficiency

As predicted, matched practice was more efficient than the combination of lecture and practice (Figure 7B). In the combined instruction condition, participants averaged 17.5 min to view the recorded lecture and complete 20 practice problems, whereas those in the matched practice condition took 15 min on average to complete the 47 practice problems, a 15% time savings, $b = -2.50$, $t(313) = -4.27$, $p = .009$. There were no significant interactions with baseline interest or confidence, $p \geq .063$.

Judgment of Learning and Confidence

On average, participants who received either type of instruction judged that they had been better prepared than those in the no instruction condition, $b = 1.24$, $d = 1.00$, $t(390) = 8.58$, $p < .001$. There was no significant difference between the two instruction conditions, $b = -0.04$, $d = -0.03$, $t(390) = -0.28$, $p = .777$, and there were no significant interactions with baseline interest or confidence, $p \geq .107$.

Confidence About the Test

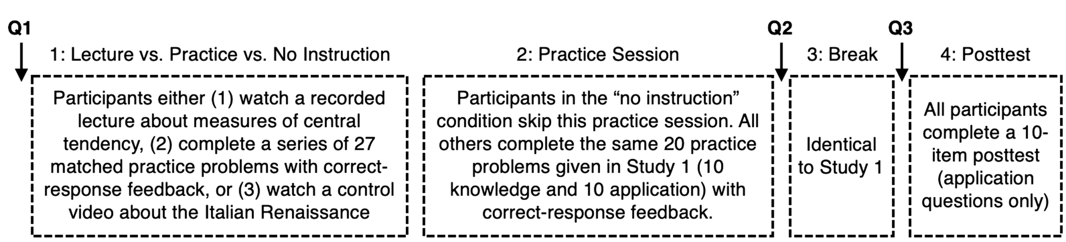
On average, participants who received either type of instruction were more confident about the upcoming test than those in the no instruction condition, $b = 0.47$, $d = 0.42$, $t(390) = 4.06$, $p < .001$. There was no significant difference between the two instruction conditions, $b = 0.06$, $d = 0.06$, $t(390) = 0.63$, $p = .527$, and no significant interactions with baseline interest or confidence, $p \geq .126$.

Distraction During the Learning Session

On average, participants who received either type of instruction reported less distraction during the learning session than those who watched the control video, $b = -1.03$, $d = -0.92$, $t(390) = -7.34$, $p < .001$, an effect that was primarily driven by participants with lower levels of initial confidence, $b = 0.42$, $t(390) = 2.48$, $p = .014$. In addition, participants who completed the matched practice

Figure 6

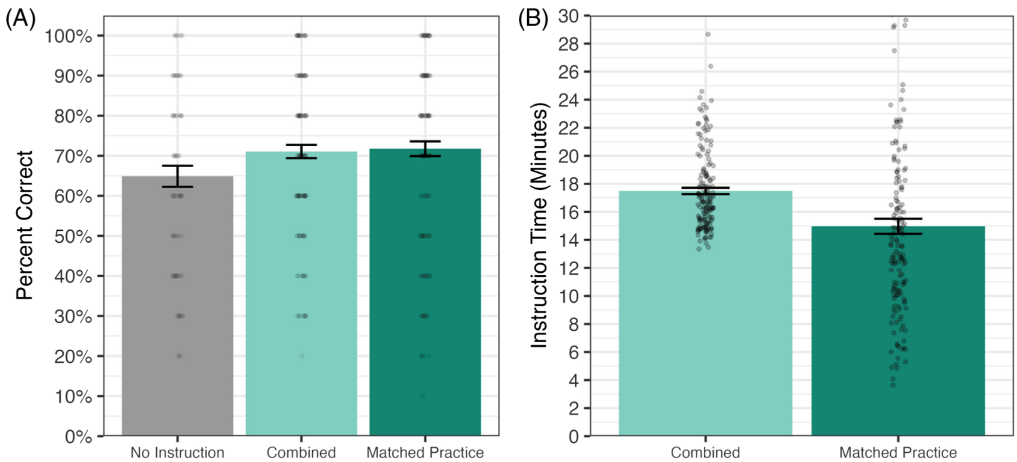
Timeline of Study 3



Note. Questionnaire 1 (Q1) assessed students' baseline interest in statistics and confidence in mathematics. Q2 assessed students' interest, beliefs about usefulness, and distraction. Q3 assessed students' judgments of learning and confidence.

Figure 7

Posttest Scores and Instruction Time, Study 2



Note. Bars represent group means for posttest scores (A) and instruction time (B); error bars show ±1 standard error of each mean. Each dot represents an individual participant, jittered on the *x*-axis to show dispersion. See the online article for the color version of this figure.

reported less distraction than those in the combined instruction condition, $b = -0.49$, $d = -0.44$, $t(390) = -4.03$, $p < .001$, an effect that was primarily driven by those with lower levels of initial interest, $b = 0.32$, $t(390) = 2.11$, $p = .036$, Figure 8A.

Interest in Measures of Central Tendency

On average, participants who received either type of instruction reported higher levels of interest in measures of central tendency than those in the control condition, $b = 0.33$, $d = 0.30$, $t(390) = 2.66$, $p = .008$. There was no significant overall difference between the two instruction conditions, $d = 0.14$, $t(390) = 1.48$, $p = .139$. However, a significant interaction between the *Practice versus Combined Instruction* contrast and baseline interest in statistics suggested that practice-based instruction was better at promoting interest than combined instruction for participants with lower levels of initial interest, $b = -0.31$, $t(390) = -2.33$, $p = .020$, Figure 8B. In addition, although the interaction between the *Practice versus Combined Instruction* contrast and baseline confidence was nonsignificant, $b = 0.24$, $t(390) = 1.76$, $p = .078$, it showed the same pattern that we observed in Study 2 and provides additional evidence that practice-based instruction may be more effective at promoting interest for students with higher levels of baseline confidence, even compared to a combination of practice and lecture, Figure 8D.

Beliefs About Usefulness

Relative to the control condition, both types of instruction promoted participants' beliefs that it is useful to know about measures of central tendency, $b = 0.32$, $d = 0.30$, $t(390) = 2.67$, $p = .008$. Furthermore, participants who received combined instruction reported beliefs about usefulness that were 0.33 *SD* higher than those who received practice-based instruction, $b = -0.33$, $d = -0.32$, $t(390) = -3.17$, $p = .002$. A nonsignificant interaction, Figure 8C, suggests that this difference was primarily driven by positive effects of combined instruction for less confident participants, $b = 0.24$, $t(390) = 1.88$, $p = .061$.

Discussion

In Study 3, we compared combined instruction (a video followed by 20 practice problems) to matched practice (29 practice problems that were matched to the video, followed by 20 additional practice problems). Although they took 15% less time to complete the instruction, participants in the matched practice condition performed just as well on a test of application questions as those who received combined instruction. In addition, the practice-based instruction was more engaging for participants with lower levels of baseline interest. For this group, the hands-on learning format held their attention during the learning session, which might have contributed to their increased interest in the material. This finding aligns with other studies that highlight the effectiveness of hands-on activities for triggering initial interest in academic content (Renninger et al., 2019).

However, Study 3 also replicated that lectures may be motivationally important for less confident students. As observed in Study 2, practice-based instruction was only helpful at promoting interest in the material for students with higher levels of confidence, who may have appreciated the challenge it provided. In addition, although confident participants reported moderately high beliefs about the usefulness of the material regardless of their instructional condition (possibly influenced by prior experience with the topic), less confident students benefitted from additional support (specifically, combined instruction that included discussion about the material's utility) to appreciate its value.

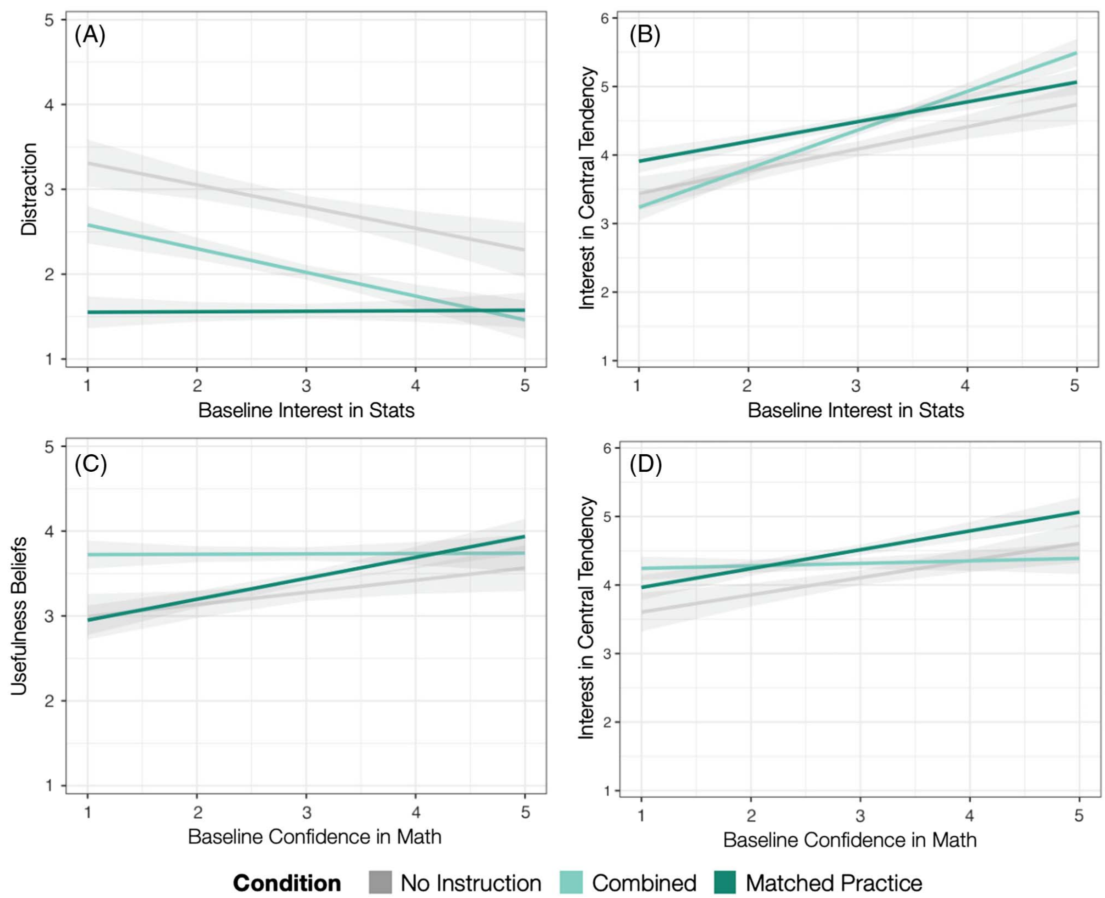
Conclusions

To grow and diversify STEM fields, it is critical to focus on the efficiency of STEM instruction. In the present research, we identified a potentially powerful lever of change to increase efficiency: Participants who were taught via practice and feedback learned more rapidly than those who completed standard instruction.

These findings, which challenge the assumption that students must learn from direct instruction before they can try things themselves and learn by doing, are consistent with the KLI framework

Figure 8

Moderated Effects of Practice Versus Combined Instruction on Interest, Distraction, and Beliefs About Usefulness



Note. (A, B) Effects of condition on distraction and interest in central tendency, moderated by baseline interest in statistics. (C, D) Effects of condition on usefulness beliefs and interest in central tendency, moderated by baseline confidence in math. See the online article for the color version of this figure.

(Koedinger et al., 2012). In this framework, lecture and feedback can both serve as instructional events, but feedback is also a metacognitively useful assessment event that helps students identify their misunderstandings, helping them more deeply attend to and process relevant information. Thus, feedback can be a more effective way to teach, and lectures can be redundant in contexts where students have access to practice opportunities and feedback. These findings are also consistent with attentional accounts of the benefits of prequestions, which suggest that questions before instruction can improve learners' attention to and processing of the correct answer, yielding more robust and durable knowledge (Carpenter & Toftness, 2017; Pan & Carpenter, 2023; Sana et al., 2020). Practice problems are prequestions for the feedback that follows, which should help learners process that feedback.

Notably, our findings about the effectiveness of practice-based instruction contrast with a recent laboratory experiment by Martella et al. (2024). Their study showed that pure lecture produced superior learning of associated pairs compared to an active learning approach. We believe this discrepancy highlights the challenges of working with the broad labels of *lecture*, which can include elements of practice testing with feedback, and *active learning*, which can refer to any activity that involves student participation.

In Martella and colleagues' study, their lecture may have been particularly effective because it included practice and feedback in the form of a guided review activity. In this activity, the instructor connected associated pairs one-by-one, pausing before they provided each answer and encouraging participants to attempt the pair themselves. Although participants were not asked to perform any active behaviors, they had the opportunity to test themselves before hearing the correct answer (i.e., feedback).

In contrast, their active learning manipulation may have been a relatively weak implementation of practice testing. Specifically, participants used a drag-and-drop interface to repeatedly group pairs of terms together. After each response, participants received feedback in the form of a sound and a green or red light. To facilitate efficient and effective learning, we propose that practice activities should test students' recall of key information, provide feedback about their knowledge, and then guide their attention to correct answers. Rather than consistently testing participants' recall of associations, the drag-and-drop activity may have allowed participants to solve the task by using superficial features (e.g., remembering the position of matches on the screen), and although the activity provided participants with feedback about their knowledge, it may not have guided their attention to correct responses.

Next Steps for Research on Learning From Practice and Feedback

Even though the present research shows that students *can* efficiently learn through practice and feedback, more work is needed to establish if and when students *should* learn in this manner. Studies 2 and 3 suggested that practice-based instruction has the potential to undermine interest for less confident students, who may be discouraged when challenged to solve problems without upfront instruction and learn from their mistakes. These findings highlight that caution may be needed when implementing "desirable difficulties" (R. A. Bjork, 1994). A more difficult learning environment is likely to improve learning as long as the difficulty is connected to processing and retrieval of relevant information (E. L. Bjork & Bjork, 2011). However, if that difficulty negatively impacts the likelihood of students' continued engagement with the material, it can become undesirable (Zepeda et al., 2020). In the case of STEM courses, interest in the material is one of the strongest predictors of long-term persistence, so classroom practices that undermine interest for struggling students could increase attrition and exacerbate inequality (Maltese & Tai, 2011; Rosenzweig et al., 2021). We propose that further research is needed to understand desirable difficulties beyond their impact on cognitive processing. In STEM courses, our work also suggests that careful design is needed to implement practice-based instruction in a way that is engaging and motivating.

In addition, it is necessary to examine the effectiveness of repeated, practice-based instruction over time, with different types of content and assessment. It may be that practice is an efficient way to make modest gains when learning the basics of a new domain, but lecture is also needed to master complex academic content. The present research suggests that a short set of practice problems can be an effective replacement for a 15-min lecture. However, can more time-consuming, repeated practice activities effectively replace engaging, full-length STEM lectures? And can practice without lecture promote deeper conceptual knowledge than we were able to teach and test in a single laboratory session (the type that would allow students to write a strong article, design their own experiment, or tackle an applied problem)? Prior classroom studies have shown that discovery-driven activities can prepare students to learn from direct instruction and that these benefits can persist over the duration of a full year (Schwartz & Martin, 2004). If both lecture and feedback can serve as instructional events, as the KLI framework suggests and as we argue in this article, we predict that well-designed practice and feedback should be effective substitutes for lectures in real STEM classrooms. To test this prediction, classroom experiments over longer periods of time are needed.

Finally, what does it mean for practice and feedback to be "well-designed?" At least three dimensions are critical for further study: how practice questions build on learners' prior knowledge, how practice and feedback can engage and motivate learners, and how feedback is provided.

Regarding prior knowledge, cognitive load theory suggests that complex problems with many unlearned, interacting elements may overwhelm working memory and be unproductive learning events (Kirschner, 2002). Therefore, for students to learn from practice and feedback, they may be best served by topics that build upon prior knowledge (Ashman et al., 2020). In the present research, this principle likely contributed to the overall success of practice-based instruction. In Studies 1 and 3, participants learned about a topic, central tendency, that was familiar to many. In Study 2, participants learned about a topic (multiple regression) that should have been much less familiar. However, we took care when creating the instructional materials to connect the new content to skills that students learned in prior algebra coursework (e.g., the equation of a line, simplifying expressions). For complex topics with many interacting, unlearned elements, upfront lectures might help students build the background knowledge that they need to learn from practice. However, we expect that scaffolded practice and feedback could be an even more efficient manner of helping novice learners build prerequisite knowledge before advancing to more complex concepts.

Regarding engagement and motivation, in Study 3, we found that the lecture, in which the instructor discussed the relevance of linear regression, boosted participants' beliefs about the content's usefulness, with larger effects for less confident participants. This finding emphasizes that, if not executed carefully, a shift toward practice problems could do away with the benefits of potentially engaging lectures. Students may be able to gain similar or greater benefits from activities that allow them to discover the value of course content as they solve and reflect about personally relevant problems (Asher et al., 2023; Bernacki & Walkington, 2018; Harackiewicz et al., 2016; Yeager et al., 2014), but such activities would need to be developed with care and research.

Regarding feedback, it will be important to study which types and features are best for replacing lectures. If, as we suggest in this article, responding to a practice question helps students attend to and learn from information that follows, this should help students learn from elaborated feedback (like we gave in Study 2) that explains *why* answers were correct or incorrect, and which should promote conceptual knowledge that helps with generalization (Mertens et al., 2022). Although participants in Studies 1 and 3 who learned from practice and feedback performed approximately as well as those who also received a lecture, the simple correct-response feedback given in these studies may have limited the performance of participants in these conditions. Thus, these studies may demonstrate a lower bound of how much one can learn from practice and feedback alone. It is also important to examine whether certain types of feedback can be motivationally helpful. For instance, warm feedback that emphasizes that struggle is common and temporary, or even a productive part of the learning process, may help keep less confident students motivated when they receive critical feedback (Fong et al., 2019).

If effective, motivating, practice-based methods of instruction can be developed to replace lectures, STEM courses may become more efficient and equitable learning opportunities.

References

Asher, M. W., & Carvalho, P. F. (2024). *Materials and code for "Practice with feedback vs. Lecture: Consequences for learning, efficiency, and motivation."* https://doi.org/10.17605/OSF.IO/3GHWV

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. Advance online publication. https://doi.org/10.1037/edu0000921

Asher, M. W., Harackiewicz, J. M., Beymer, P. N., Hecht, C. A., Lamont, L. B., Else-Quest, N. M., Priniski, S. J., Thoman, D. B., Hyde, J. S., & Smith, J. L. (2023). Utility-value intervention promotes persistence and diversity in STEM. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(19), Article e2300463120. https://doi.org/10.1073/pnas.2300463120

Ashman, G., Kalyuga, S., & Sweller, J. (2020). Problem-solving or explicit instruction: Which should go first when element interactivity is high? *Educational Psychology Review*, *32*(1), 229–247. https://doi.org/10.1007/s10648-019-09500-5

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research*, *70*(2), 181–214. https://doi.org/10.3102/00346543070002181

Bandura, A. (1986). *Social foundations of thought and action*. Prentice-Hall.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

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

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. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). MIT Press. https://doi.org/10.7551/mitpress/4561.001.0001

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

Canning, E. A., & Harackiewicz, J. M. (2015). Teach it, don't preach it: The differential effects of directly-communicated and self-generated utility-value information. *Motivation Science*, *1*(1), 47–71. https://doi.org/10.1037/mot0000015

Carpenter, S. K., Pan, S. C., & Butler, A. C. (2022). The science of effective learning with spacing and retrieval practice. *Nature Reviews Psychology*, *1*(9), 496–511. https://doi.org/10.1038/s44159-022-00089-1

Carpenter, S. K., & Toftness, A. R. (2017). The effect of prequestions on learning from video presentations. *Journal of Applied Research in Memory and Cognition*, *6*(1), 104–109. https://doi.org/10.1016/j.jarmac.2016.07.014

Carvalho, P. F., McLaughlin, E. A., & Koedinger, K. R. (2022). Varied practice testing is associated with better learning outcomes in self-regulated online learning. *Journal of Educational Psychology*, *114*(8), 1723–1742. https://doi.org/10.1037/edu0000754

Chen, X. (2013). *STEM attrition: College students' paths into and out of STEM fields*. National Center for Education Statistics, Education Sciences, U.S. Department of Education. https://nces.ed.gov/pubs2014/2014001rev.pdf

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*, Article 101859. https://doi.org/10.1016/j.cedpsych.2020.101859

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363–406. https://doi.org/10.1037/0033-295X.100.3.363

Fong, C. J., Patall, E. A., Vasquez, A. C., & Stautberg, S. (2019). A meta-analysis of negative feedback on intrinsic motivation. *Educational Psychology Review*, *31*(1), 121–162. https://doi.org/10.1007/s10648-018-9446-6

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

Gaspard, H., Dicke, A.-L., Flunger, B., Brisson, B. M., Häfner, I., Nagengast, B., & Trautwein, U. (2015). Fostering adolescents' value beliefs for mathematics with a relevance intervention in the classroom. *Developmental Psychology*, *51*(9), 1226–1240. https://doi.org/10.1037/dev0000028

Harackiewicz, J. M., Canning, E. A., Tibbetts, Y., Priniski, S. J., & Hyde, J. S. (2016). Closing achievement gaps with a utility-value intervention: Disentangling race and social class. *Journal of Personality and Social Psychology*, *111*(5), 745–765. https://doi.org/10.1037/pspp0000075

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

Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, *53*(3), 983–997. https://doi.org/10.2307/2533558

Kirk-Johnson, A., Galla, B. M., & Fraundorf, S. H. (2019). Perceiving effort as poor learning: The misinterpreted-effort hypothesis of how experienced effort and perceived learning relate to study strategy choice. *Cognitive Psychology*, *115*, Article 101237. https://doi.org/10.1016/j.cogpsych.2019.101237

Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learning and Instruction*, *12*(1), 1–10. https://doi.org/10.1016/S0959-4752(01)00014-7

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

Koedinger, K. R., Carvalho, P. F., Liu, R., & McLaughlin, E. A. (2023). An astonishing regularity in student learning rate. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(13), Article e2221311120. https://doi.org/10.1073/pnas.2221311120

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

Koedinger, K. R., Kim, J., Jia, J. Z., McLaughlin, E. A., & Bier, N. L. (2015). Learning is not a spectator sport: Doing is better than watching for learning from a MOOC. *Proceedings of the second (2015) ACM conference on learning @ scale* (pp. 111–120). https://doi.org/10.1145/2724660.2724681

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

Laursen, S. (2019). *Levers for change: An assessment of progress on changing STEM instruction*. American Association of the Advancement of Science. https://www.aaas.org/sites/default/files/2019-07/levers-for-change-WEB100_2019.pdf

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

Macnamara, B. N., Hambrick, D. Z., & Oswald, F. L. (2014). Deliberate practice and performance in music, games, sports, education, and professions: A meta-analysis. *Psychological Science*, *25*(8), 1608–1618. https://doi.org/10.1177/0956797614535810

Maltese, A. V., & Tai, R. H. (2011). Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among U.S. students. *Science Education*, *95*(5), 877–907. https://doi.org/10.1002/sce.20441

Martella, A. M., Schneider, D. W., O'Day, G. M., & Karpicke, J. D. (2024). Investigating the intensity and integration of active learning and lecture. *Journal of Applied Research in Memory and Cognition*, *13*(3), 354–369. https://doi.org/10.1037/mac0000160

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

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), Article 97. https://doi.org/10.1007/s10648-023-09814-5

R Core Team. (2023). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

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

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*, Article 100260. https://doi.org/10.1016/j.lcsi.2018.11.007

Roediger, H. L., III, & 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

Rosenzweig, E. Q., Harackiewicz, J. M., Hecht, C. A., Priniski, S. J., Canning, E. A., Tibbetts, Y., Asher, M. W., & Hyde, J. S. (2021). College students' reasons for leaving biomedical fields: Disenchantment with biomedicine or attraction to other fields? *Journal of Educational Psychology*, *113*(2), 351–369. https://doi.org/10.1037/edu0000456

Rothkopf, E. Z. (1968). Textual constraint as function of repeated inspection. *Journal of Educational Psychology*, *59*(1), 20–25. https://doi.org/10.1037/h0025378

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys* (1st ed.). Wiley. https://doi.org/10.1002/9780470316696

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

Sana, F., Forrin, N. D., Sharma, M., Dubljevic, T., Ho, P., Jalil, E., & Kim, J. A. (2020). Optimizing the efficacy of learning objectives through pretests. *CBE Life Sciences Education*, *19*(3), Article ar43. https://doi.org/10.1187/cbe.19-11-0257

Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction*, *22*(2), 129–184. https://doi.org/10.1207/s1532690xci2202_1

Stains, M., Harshman, J., Barker, M. K., Chasteen, S. V., Cole, R., DeChenne-Peters, S. E., Eagan, M. K., Jr., 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., … Young, A. M. (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

Tauber, S. K., Dunlosky, J., Rawson, K. A., Rhodes, M. G., & Sitzman, D. M. (2013). General knowledge norms: Updated and expanded from the Nelson and Narens (1980) norms. *Behavior Research Methods*, *45*(4), 1115–1143. https://doi.org/10.3758/s13428-012-0307-9

Yeager, D. S., Henderson, M. D., Paunesku, D., Walton, G. M., D'Mello, S., Spitzer, B. J., & Duckworth, A. L. (2014). Boring but important: A self-transcendent purpose for learning fosters academic self-regulation. *Journal of Personality and Social Psychology*, *107*(4), 559–580. https://doi.org/10.1037/a0037637

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

Received January 29, 2024
Revision received September 11, 2024
Accepted October 21, 2024 ■