Title

Students Can Learn More Efficiently When Lectures Are Replaced with Practice Opportunities and Feedback

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors
Asher, Michael W
Sana, Faria
Koedinger, Kenneth R
et al.

Publication Date
2024

Peer reviewed

Students Can Learn More Efficiently When Lectures Are Replaced with Practice Opportunities and Feedback

Michael W. Asher (masher@andrew.cmu.edu)
Human Computer Interaction Institute, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213

Faria Sana (fsana@athabascau.ca)
Center for Social Sciences, Athabasca University
10011 109 Street Northwest, #1200, Edmonton, Alberta T5J 3S8, Canada.

Kenneth R. Koedinger (koedinger@cmu.edu)
Paulo F. Carvalho (pcarvalh@andrew.cmu.edu)
Human Computer Interaction Institute, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213

Abstract

Many college students drop out of STEM majors after struggling in gateway courses, in part because these courses have large time demands. The risk of attrition is higher for those from financially disadvantaged backgrounds who often work to pay for college, making such time commitments unfeasible. In two laboratory experiments with different topics (central tendency and linear regression), we identified a promising approach to increase the efficiency of STEM instruction. When we removed instructional videos and taught participants exclusively with practice and feedback, they learned 2-3 times 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, more equitable learning environments.

Keywords: active learning; practice testing; STEM education

Introduction

Every year many aspiring STEM majors, and particularly those from financially-disadvantaged backgrounds, change their majors or even drop out of college after struggling in introductory math and science courses (Chen, 2013). One cause of this problem is that STEM courses place large demands on students' time (Seymour & Hunter, 2019). Although the time demands of a STEM-heavy course load may not be an obstacle for students with flexible schedules, they place a particular burden on financially disadvantaged students who often work to pay for college (Carnevale & Smith, 2018). To help lower-income students succeed and persist in STEM, instructors must strive to make their courses efficient learning opportunities.

Fortunately, STEM courses may have substantial room to improve how they allocate students' time. STEM instructors often invest large amounts of class time lecturing about course content (Stains et al., 2018), and students then practice the same content 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 yield more efficient learning?

The importance of practice is well established. Research on topics like the testing effect, active learning, and deliberate practice show 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; 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. Specifically, there's little experimental evidence about whether lecture is needed if participants are already learning from 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. For instance, repetition usually has a positive effect on learning and retention, even with passive instructional approaches (Rothkopf, 1968). Additionally, when an instructor explains a topic or demonstrates a skill during a lecture before students practice it, students may be more likely to practice (and thereby reinforce) correct responses (Carvalho et al., 2022).

However, it is also possible that lectures are redundant and inefficient if students already have access to practice opportunities and high-quality feedback. The Knowledge-Learning-Instruction 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. Like learning events, changes to knowledge can't be directly observed. Instead, they are inferred with assessment events (e.g., homework assignments, essays, tests, discussions, etc.).

When students passively learn via lecture, they complete instructional events. However, when students attempt to solve problems and receive feedback, they complete both

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

instructional events and assessment events. Critically, this combination of instruction and assessment may lead 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 (Bjork et al., 2013). 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.

Evidence for this process can be found in studies of "pretests" or "prequestions" in which instructional events like written passages or lectures are preceded by questions about their contents. For example, Carpenter & Toftness (2017), randomly assigned students to complete a set of short-answer practice questions, or not, before they watched a recorded history lecture. No immediate feedback was provided about the correctness of students' responses; instead, students inferred this information from the lecture. On an end-of-session test, participants who completed prequestions performed a full standard deviation better than participants in the control group, suggesting that the prequestioned group was able to more effectively attend to the lecture. Analyses of student behavior in online courses also provide evidence that lecture may not be needed when students have the opportunity to 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 investigated this hypothesis.

Study 1

We conducted Study 1 to (a) establish if participants could learn from practice-based instruction (consisting of practice with feedback, but no lecture), (b) compare practice-based instruction to the standard approach in STEM courses: a combination of lecture and practice with feedback, and (c) assess how the different forms of instruction affected participants' subjective judgments of learning. In this study, participants learned about a statistics topic and then took a test. To test our research questions, we randomly assigned participants to one of four conditions: no instruction, practice with feedback only, lecture only, or combined instruction.

We hypothesized that practice with feedback would be an effective instructional event, such that students 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 (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).

Participants

A total of 132 participants were recruited through Prolific who consented to participate and completed the study. 97 (79%) self-reported their ethnicity as White, 14 (11%) as Black, 10 (8%) as Asian, 4 (3%) as multiracial, and 2 (2%) as belonging to another group. Two participants (2%) chose not to report their ethnicity. 79 participants (60%) identified as women and 50 (28%) as men. Three participants chose not to report their sex (2%). The average age of participants was 41.6 years. The study lasted approximately 30 minutes, and participants were paid $4.80 for their participation.

Procedure

The study contained four sections: an instructional video, practice, a five-minute break, and then a test. During the video portion, participants who were randomized to the "lecture" and "combined" conditions watched a video about measures of central tendency. All other participants watched a control video, 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 five minutes. Finally, all participants proceeded to a survey and then a posttest about measures of central tendency.

Materials

For the video, we used an 11.4-minute 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-minute CrashCourse video about the Italian Renaissance. To create practice and assessment materials for the lesson, we developed 20 "knowledge" questions that tested facts that were covered in the video (e.g., definitions), and 20 "application" questions that required participants to apply the information. Because individuals best learn to generalize when they're exposed to varied input (Raviv et al., 2022), we wanted to examine if participants would need the variety of combined instruction to succeed on 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. All questions were 4-item multiple choice, and the order of questions was randomized for each participant. The trivia questions given during the five-minute break were developed and normed by Tauber and colleagues (2013).

In addition, we tracked instruction time for each participant (i.e., the time they spent on the video and/or practice sections). We inserted a single item after instruction, assessing participants' self-reported judgments of their own learning ("The instruction I just received prepared me well to answer questions about measures of central tendency"), adapted from Koriat & Ackerman (2010). Participants responded to this item using a slider that ran from "0: Strongly Disagree" to "100: Strongly Agree." Following a five-minute break and posttest, 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 instruction 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.

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 datasets, each with different plausible estimates for students' initial knowledge in all conditions. 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 (Rubin, 1987).

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 hypothesis about performance, we calculated each participant's scores on memory and generalization questions on the posttest, and then fit a series of three linear mixed effects models, each 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 vs. Knowledge contrast, which indicated whether each score was for application (.5) or knowledge (-.5) questions, a Version contrast to control for whether participants were assigned to version "A" (.5) or "B" (-.5) of the problem set, the two-way interactions between the Application vs. Knowledge contrast and the dummy coded condition contrasts, a by-participant random intercept, and a by-participant random slope for the Application vs. Knowledge contrast. p values were adjusted for multiple comparisons (Benjamini & Hochberg, 1995).

Judgments of learning, a between-subjects variable, were 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 order.

Performance. The left panel of Figure 1A 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, $d = 1.03$, $F(1, 127) = 16.95$, $p < .001$, 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, $d = .29$, $F(1, 127) = 1.17$, $p = .266$. 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, $d = .43$, $F(1, 127) = 2.92$, $p = .135$.

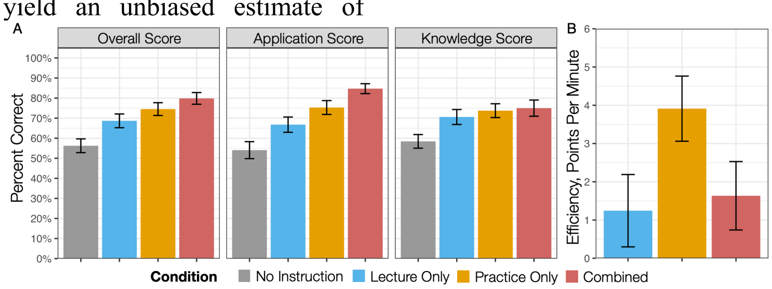
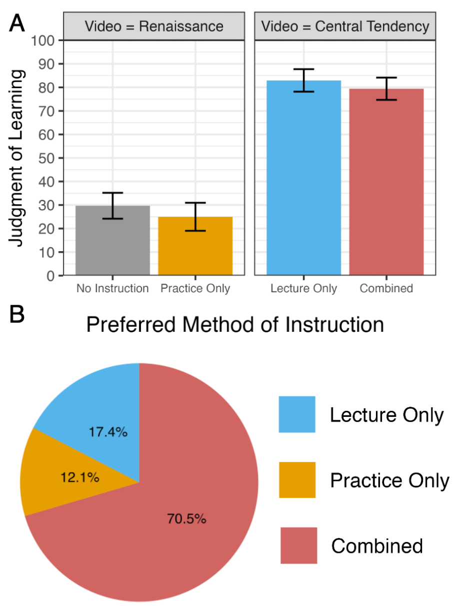


Figure 1: Posttest Scores and Efficiency of Learning

The right two panels of Figure 1A display average performance on application and knowledge items. Two significant interactions with the Application vs. Knowledge contrast suggested that combined instruction was particularly good at promoting application (vs. knowledge), relative to no instruction, $b = .14$, $F(1, 128) = 10.13$, $p = .007$, and lecture-based instruction, $b = .14$, $F(1, 128) = 9.64$, $p = .007$. No other interactions with the Application contrast were significant, $p \geq .117$.

Efficiency. Figure 1B 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, 2.4x more efficiently than those in the combined condition (1.63 points per minute), $t(87.25) = -3.58$, $p = .001$, and more than 3x more efficiently than those in the lecture only condition (1.24 points per minute), $t(89.99) = -4.11$, $p < .001$.

Figure 2: Judgments of Learning and Preferences



Judgments of Learning and Instruction Preferences. Figure 2A 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, $d = -.15$, $t(117) = -.60$, $p = .662$. Similarly, participants in the combined condition reported similar judgments of learning to those in the lecture only condition, $d = -.08$, $t(117) = -.32$, $p = .747$, 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 3 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 2B.

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, this benefit was offset by a massive loss in efficiency. Practice was the most efficient form of instruction by a wide margin: participants in the practice-only condition learned more than twice as quickly as those in the combined condition, and more than 3x faster than those in the lecture only condition. This finding suggests that if instructors want to improve their students' ability to generalize, they may be best served by having their students practice a variety of topics and skills and then providing feedback, which would allow students to experience a wider range of input than they would receive with lecture in the same amount of time.

However, we also observed that participants judged that practice 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 learning via practice and feedback can be good for learning, it may undermine student motivation. We tested this possibility in Study 2.

Study 2

In Study 2, we set out to replicate and extend the findings of Study 1 in a new context, with new materials. In addition to focusing on learning and efficiency of instruction as outcomes, we examined how practice-based instruction affects motivation, and we did so in a context that more closely resembles a college-level STEM course. We designed an experimental paradigm in which college-student participants learned the basics of multiple regression, as they would in an introductory statistics course, either with a video or with practice problems and feedback. To increase the potential effectiveness of the video, 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 to cover the same information as the lecture, in the same order. After instruction, we gave all participants an application-heavy test of items that were new for all participants.

We predicted that participants would learn more from practice than the video 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**.**

Participants

Undergraduate participants were recruited 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, 4 (1%) as Middle Eastern, 3 (1%) as Indigenous, and 1 (< 1%) as belonging to another group. 212 participants (63%) identified as women and 123 (36%) as men. Two participants identified as non-binary (<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.

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-minute video that we recorded, or they completed a series of 16 practice problems. After each problem, participants received feedback. The video was built around a series of two worked examples, which occupied 51% of the video. In addition, the instructor spent 32% of the video on definitions and explanations, 12% on transitions between topics, and 5% on a recap. Unlike the prior two studies, there was no control video in the practice-only condition to facilitate an unconfounded test of how the two manipulations affected students' interest. All participants worked through the practice problems in the same order.

Finally, participants in both conditions completed a 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. Parts 2-4 were each built around a different equation or graph (e.g., a scatter plot of the relationship between gas prices and traffic fatalities) and contained short-answer application questions.

Measures

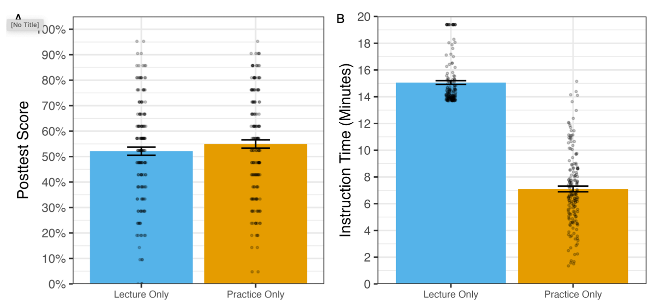
Baseline interest in statistics was measured with three items (e.g. "How interesting do you find statistics?", α = .93), as was baseline confidence in math (e.g., "How good are you at math?", α = .93). On the outcome questionnaire, interest in statistics was measured with 12-items (e.g. "How interesting do you find linear regression?", "How much would you enjoy learning more about statistics in the future", α = .96). and we measured confidence in regression with three items (e.g., "How well do you think you would do in a regression course?", α = .93). These scales were adapted from Linnenbrink-Garcia et al. (2010) and Durik et al. (2015). In addition, participants received an overall score on the posttest; the test's design did not allow for separate analyses of memory and generalization questions. Because Study 2 lacked counterbalanced pre 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 vs. 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 explore if less-interested students would prefer the recorded video or practice.

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, $d = .12$) and non-significant, $t(332) = 1.22$, $p = .225$, Figure 3A. There were no significant interactions with baseline interest or confidence, $p \geq .071$.

Figure 3: Posttest Scores and Efficiency of Learning

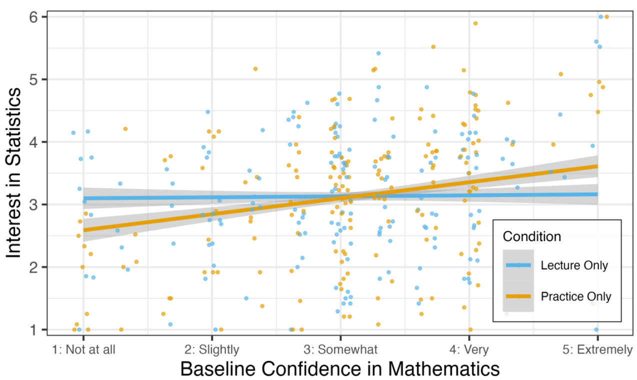


Efficiency. Again, practice was a much more efficient form of instruction than lecture. Whereas participants took 15.1 minutes on average to watch the video, they averaged only 6.7 minutes to complete the practice problems, $t(332) = -31.30$, $p < .001$, Figure 3B. Assuming that participants in both conditions learned a similar amount, this translates to more than a 2.25x 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 well for the test, although the effect on this outcome was in the predicted direction: judgments of learning were .15 SD lower in the practice condition than in the lecture condition, $d = -.15$, $t(332) = -1.43$, $p = .154$. There were no significant interactions with baseline interest or confidence, $p \geq .236$.

Interest in Statistics. In the two conditions, after instruction participants reported similar levels of interest in statistics on average, d = -.02, t(332) = -0.18, p = .860. However, a significant Practice vs. 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 = .23, t(332) = 2.13, p = .034, Figure 4. There was no significant interaction with baseline interest in statistics, p = .678.

Figure 4: Interest in Statistics by Condition and Baseline Confidence in Mathematics



Discussion

After participants learned either from a lecture or a closely-matched set of practice problems with feedback, they performed similarly well on an application-heavy test that required them to generalize what they had learned to solve novel problems. In addition, Study 2 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 suggests that practice with feedback can remain efficient even when it covers all the material in a lecture, and 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.

Conclusions

To grow and diversify STEM fields, it is critical to focus on the efficiency of instruction in introductory STEM courses. In the present research we identified a potentially powerful lever of change to do this: participants who were taught via practice and feedback learned 2-3 times as rapidly as those who completed standard instruction (i.e., practice and feedback, plus lecture).

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 Knowledge-Learning-Instruction framework (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 to correct them. Thus, feedback can be a more effective way to teach, and lectures should be redundant in contexts where students have access to practice opportunities and high-quality 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; Sana et al., 2020).

Yet, even though this 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. Study 2 showed 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" (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 (Bjork & Bjork, 2011), as is the case in our practice-only condition. However, if that difficulty negatively impacts the likelihood of students' continued engagement with the material, it can become undesirable. 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. It may be that practice with feedback 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 brief instructional video, but can more time-consuming, repeated practice activities effectively replace engaging, full-length STEM lectures? To answer these questions and begin collecting evidence on the actual impact of practice-based instruction, it will be necessary to move beyond the laboratory and conduct field experiments in STEM classrooms.

Acknowledgements

This research is supported in part by funding from the Walton Family Foundation, NSF (#2301130), and the SSHRC of Canada (#435-2021-0426).

References

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.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In *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. *Metacognition: Knowing about Knowing.*, 185–205. 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

Carnevale, A., & Smith, N. (2018). *Balancing work and learning: Implications for low-income students.* Georgetown University Center on Education and the Workforce. https://cew.georgetown.edu/cew-reports/LearnAndEarn/#full-report

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*, 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

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*, *111*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

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*, 101237. https://doi.org/10.1016/j.cogpsych.2019.101237

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*, *120*(13), 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*, 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

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

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

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

Rosenzweig, E. Q., Hecht, C. A., Priniski, S. J., Canning, E. A., Asher, M. W., Tibbetts, Y., Hyde, J. S., & Harackiewicz, J. M. (2021). Inside the STEM pipeline: Changes in students' biomedical career plans across the college years. *Science Advances*, *7*(18), eabe0985. https://doi.org/10.1126/sciadv.abe0985

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

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), ar43. https://doi.org/10.1187/cbe.19-11-0257

Seymour, E., & Hunter, A.-B. (Eds.). (2019). *Talking about leaving revisited: Persistence, relocation, and loss in undergraduate STEM education*. Springer International Publishing. https://doi.org/10.1007/978-3-030-25304-2

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., … Young, A. M. (2018). Anatomy of STEM teaching in North American universities. *Science*, *359*(6383), 1468–1470. https://doi.org/10.1126/science.aap8892

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