



The Value of Using Tests in Education as Tools for Learning—Not Just for Assessment

Dillon H. Murphy¹ · Jeri L. Little² · Elizabeth L. Bjork¹

Accepted: 28 August 2023 / Published online: 8 September 2023
© The Author(s) 2023

Abstract

Although students tend to dislike exams, tests—broadly defined in the present commentary as opportunities to practice retrieving to-be-learned information—can function as one of the most powerful learning tools. However, tests have a variety of attributes that affect their efficacy as a learning tool. For example, tests can have high and low stakes (i.e., the proportion of a student’s grade the exam is worth), vary in frequency, cover different ranges of course content (e.g., cumulative versus non-cumulative exams), appear in many forms (e.g., multiple-choice versus short answer), and occur before or after the presentation of what is to be learned. In this commentary, we discuss how these different approaches to test design can impact the ability of tests to enhance learning and how their use as instruments of learning—not just means of assessment—can benefit long-term learning. We suggest that instructors use frequent, low-stakes, cumulative exams and a variety of test formats (e.g., cued recall, multiple-choice, and true/false) and give students exams both prior to learning and following the presentation of the to-be-learned material.

Keywords Testing · Learning · Spacing · Generation · Desirable difficulties

Over the first two decades of life, many of us spend a huge portion of our time in school as students. As instructors, most of us strive to make these years as effective as possible for students by utilizing teaching and assessment techniques typically considered to be the best. However, in the present commentary, we propose that despite having this admirable goal of doing our best to optimize the quality of learning achieved by our students, we often do not implement, or at least not to the

This article is part of the Topical Collection on Test-Enhanced Learning and Testing in Education: Contemporary Perspectives and Insights

✉ Dillon H. Murphy
dmurphy8@ucla.edu

¹ Department of Psychology, University of California, Los Angeles, CA, USA

² California State University, East Bay, Hayward, CA, USA

degree that we should, the most effective techniques to enhance our students' learning, particularly in terms of long-term retention and transfer. Specifically, we have come to primarily use tests as a mechanism for assessment and often overlook their use as a powerful tool for learning.

Tests: from Assessing Learning to Promote Learning

In educational settings, learning is a multifaceted process involving the acquisition, retention, and application of knowledge and skills. It encompasses not only the immediate gains in performance that can be observed during or shortly after a learning activity but also the more enduring changes in memory and understanding that lead to enhanced long-term retention and transfer of knowledge. While immediate gains in performance may give the appearance of effective learning, they can sometimes be deceiving, as they may represent only superficial improvements without true comprehension or retention. Actual learning, on the other hand, is characterized by the ability to retrieve and apply the learned information over time and in different contexts. That is, evidence of actual learning can be seen when students demonstrate the ability to recall and apply the learned material on subsequent assessments or in real-world situations, even after a delay. Additionally, instructors can look for signs of deep understanding and the ability to make connections between different concepts, indicating that students have achieved meaningful learning of the material rather than just rote memorization of it.

To assess learning, instructors can monitor students' progress over time, evaluate their performance on different types of assessments (e.g., quizzes, exams, problem-solving tasks), and provide opportunities for students to demonstrate their understanding in varied contexts. The types of assessments instructors use to see how much their students know can do more than just assess their knowledge—they can also help them learn. Additionally, feedback on assessments can play a crucial role in identifying areas for improvement and guiding students toward more effective study strategies. By carefully observing students' learning outcomes and adjusting instructional methods based on this information, instructors can create a supportive learning environment that fosters meaningful, long-lasting, and transferable learning.

The discussion above characterizes two prominent views of the role of assessment in education: summative assessment and formative assessment. Summative assessment is often associated with traditional testing methods and is used to measure overall achievement and learning outcomes at the end of a specific instructional period or course. Summative assessment is typically used to assign grades and make judgments about students' performance. However, as educators, we advocate for going beyond the traditional use of tests solely for summative purposes. Instead, we propose utilizing tests as formative assessments as well. Formative assessment involves the use of ongoing, low-stakes evaluation methods, such as practice tests and quizzes, to inform both students and teachers about the student's current level of understanding and knowledge. These assessments are intended to guide and shape the learning process, allowing students to identify areas of weakness and instructors to tailor their teaching methods accordingly. By embracing the dual role of tests as

both summative and formative evaluation tools, we can enhance students' learning experiences and promote more effective long-term retention and understanding of course content.

Formative assessment has traditionally been used to convey the way testing indirectly promotes learning (by improving future learning), but testing can also play a direct role in enhancing learning through retrieval processes. When students engage in testing, they actively retrieve information from memory, reinforcing existing retrieval routes and establishing new ones (Bjork, 1975; Carrier & Pashler, 1992; McDaniel & Masson, 1985). This process strengthens memory and facilitates better long-term retention of the material.

In this commentary, we adopt a broad definition of testing, focusing on its use as a tool for learning and encompassing both formal and informal activities that prompt students to answer questions related to course content. This definition includes traditional formal assessments such as quizzes and exams, but it also extends to other question-answering activities such as responding to polling questions or participating in review games, which may have lower stakes and be less formal. Our main goal is to encourage instructors to incorporate activities that prompt students to actively retrieve information from memory, as this process has been shown to enhance learning and long-term retention (e.g., Bjork, 1975). These activities can take different forms, such as low-stakes quizzes, clicker questions, or review games, but the underlying principle is the same: engaging students in retrieval practice to strengthen their memory representations and promote deeper learning.

How Can Testing Act as a Desirable Difficulty?

More broadly, using tests as a tool for learning represents a desirable difficulty (e.g., Bjork & Bjork, 2014, 2022; Karpicke, 2017). These learning strategies create challenges for learners, which may initially make it more difficult to perform correctly and thus appear to slow down the learning process. However, these difficulties ultimately result in the type of learning that is highly desirable: learning that is both long-lasting and transferable. Examples of such desirable difficulties include (a) spaced or distributed practice (versus blocked or massed practice; Bjork & Allen, 1970; Cepeda et al., 2006; Greene, 2008; Karpicke & Bauernschmidt, 2011; Murphy et al., 2022); (b) contextual variation (that is, changing the conditions of practice rather than keeping them constant and predictable (Imundo et al., 2021; Smith et al., 1978)); (c) interleaving (varying the topics being studied rather than studying only one over and over again before moving on to the next one (e.g., Kornell & Bjork, 2008)); and (d) testing or retrieval practice (DeWinstanley & Bjork, 2004; Halamish & Bjork, 2011; Roediger & Karpicke, 2006a).

It is important to note that desirable difficulties are not solely defined by their level of difficulty but by their ability to induce the type of cognitive processing that enhances learning. While some learning strategies may be challenging, that characteristic alone does not enable them to enhance learning. Rather, it is whether the difficulties or challenges they present lead the learner to engage in the type of cognitive processes that produce improved retention and understanding. The key lies

in the processes induced during the learning or study experience rather than the perceived difficulty of that activity itself. The term “desirable difficulties” serves to remind both instructors and students that encountering challenges during the learning activity, even when doing so may appear to slow down one’s performance gains, should not be equated with the production of poor learning outcomes. Instead, the focus should be on identifying whether the difficulties being encountered during that learning activity are leading the student to engage in effective learning processes.

Rather than focusing on making tests more difficult, instructors can better serve their students by ensuring that the students are engaging in such beneficial learning strategies during the testing experience. The goal is to design assessments that prompt active retrieval, encourage critical thinking, and foster deep engagement with the material. Instructors can use a variety of testing formats, such as multiple-choice questions with competitive alternatives, cued-recall questions, or collaborative testing, to promote engagement in these desirable learning processes. By understanding the underlying mechanisms through which testing improves learning, instructors can strategically employ testing as a powerful tool for enhancing retention, understanding, and transfer of knowledge.

What Factors Should Be Considered When Administering Tests?

When designing or administering tests, instructors make several decisions that can impact the effectiveness of the test as a tool for learning. One key decision is the number of tests and the subsequent stakes of each test (i.e., the proportion of a student’s final course grade that will be tied to their performance on each exam). For instance, having just one or two high-stakes exams (e.g., the popular use of having only a mid-term and a final) compared to including many lower-stakes exams and/or quizzes (more frequent testing can result in each test being worth less in terms of grade percentage) may be less effective at creating long-lasting learning. Another key decision for instructors to make is the range of course content that each test will cover (e.g., whether to use cumulative or non-cumulative exams). Additionally, the test format is an important decision to make, as tests can appear in many forms (e.g., multiple-choice vs. short answer), and these different forms can vary widely in their ability to function as a tool for learning.

Each of these different decisions or approaches to test design can impact the quality of learning that students will achieve (i.e., whether it will be learning that remains accessible and transferable for the long term or becomes quickly inaccessible or forgotten) and thus needs to be carefully considered by instructors. Additionally, there may be unusual approaches to testing that can enhance learning, such as using tests prior to learning (i.e., as pretests). Moreover, incorporating more competitive alternatives (i.e., those that are plausible enough to be seriously considered) into multiple-choice tests, thereby causing students to engage in more retrieval processes as opposed to recognition processes to select the correct alternative, may lead to greater retention and understanding of the tested concepts. Finally, despite the availability of such effective testing practices, these techniques may not be utilized frequently enough by instructors as part of their in-class activities or by students in

their independent study strategies (e.g., the productiveness of students' self-directed study efforts would almost certainly be enhanced by incorporating more self-testing as part of their efforts to learn outside of the classroom). We discuss each of these key decisions and their potential consequences for learning in more detail in the remainder of this commentary with the hope of making a compelling case for how a greater use of testing as a tool for learning rather than just as a means of assessment can be a way to enrich the learning of our students both in and out of the formal classroom setting.

How Can Testing Indirectly and Directly Improve Learning?

Alternatives to desirable difficulties like restudying to-be-remembered information rather than engaging in retrieval practice tend to have the appearance of speeding up learning, which is probably one of the reasons they are so widely used in instruction. However, such gains typically only represent superficial improvements in performance rather than increases in actual learning, and these improvements are not likely to last or to be transferable (e.g., Roediger & Karpicke, 2006b; Rohrer & Taylor, 2007). In contrast, introducing desirable difficulties into one's instructional practices—because they do challenge the learner—can sometimes slow down one's apparent gain in performance and thus be incorrectly interpreted as slowing down the learning process. Engaging in the use of such desirable difficulties, however, leads to learning that will be both long-lasting and transferable. Unfortunately, this contrast in immediate performance gains (which is something we can readily observe) versus actual learning (which can only be inferred or measured at a delay) can frequently lead both students and instructors to be tricked into preferring poorer methods of studying or teaching over better, more effective methods.

Testing can improve learning through two distinct routes, both of which are essential for a comprehensive understanding of its impact. The first route, often associated with “formative evaluation,” involves the indirect benefits of testing on learning. Indirect benefits include giving students a better idea of what they do or do not know, so they can plan their future study efforts more effectively (see Rhodes, 2016 for a review on how learners metacognitively monitor their learning). More specifically, students can better monitor their learning when being tested (see Narens et al., 2008) because tests reveal what information is accessible and what information they are unable to access (e.g., Little & McDaniel, 2015).

As a result, frequent testing can lead to more effective studying, whereby students spend less time studying already-mastered concepts and more time studying yet-to-be-learned material (Dunlosky & Hertzog, 1998). However, we must make our students understand that when they self-test and get things wrong, they are not failing; rather, they are identifying what they need to study more of and, thus, creating an opportunity for successful learning of that specific material. That is, we need to help our students understand that not knowing the answer to a given question does not represent a failure or something bad on their part; rather, they should view such occurrences as positive events because they create opportunities for new and effective learning.

The second route, which is not explicitly captured in either formative or summative evaluation, pertains to the direct impact of correctly recalling information during the testing process. When students successfully retrieve information from memory, the act of recalling itself strengthens and modifies the representation of that information in their memory. This process, known as retrieval practice or the testing effect, leads to improved long-term retention and the creation of more robust retrieval routes for future access (Bjork, 1975; Carrier & Pashler, 1992; McDaniel & Masson, 1985). By repeatedly testing their knowledge, students consolidate the learned material in their memory and enhance the accessibility of that information over time and in a variety of contexts.

Both routes highlight the unique benefits of incorporating testing as a powerful tool for learning enhancement. While formative evaluation captures the indirect benefits of testing, the direct impact of successful recall and retrieval practice is equally crucial for fostering durable learning outcomes. By recognizing the dual role of testing in both informing and reinforcing learning, educators can strategically design assessment practices that go beyond mere evaluation and truly optimize the learning process.

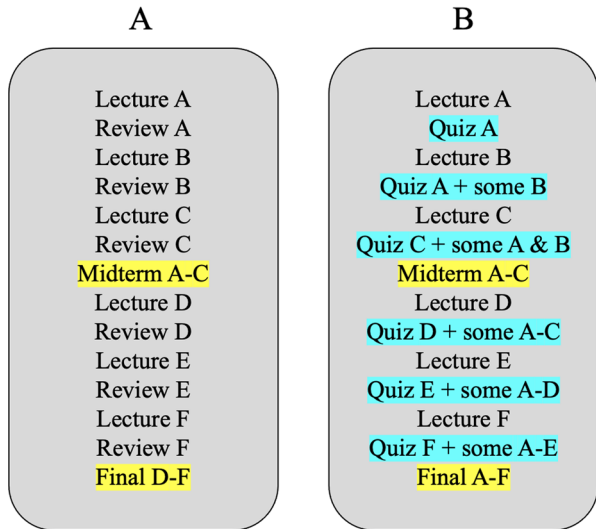
How Often Should We Give Tests?

As educators, we should shift our mindset from viewing assessments solely as tests to measure learning (usually at the end of blocks of instruction) to a broader perspective where assessment becomes a powerful tool for enhancing learning (see also Roediger et al., 2011 for the benefits of testing). Doing so means incorporating assessments or testing more frequently throughout the instructional process. Although we should continue to give exams to measure what has been learned after a period of instruction, we should stop thinking of that occasion as the only or main time to employ testing with respect to the learning of that material. We should capitalize on the power of testing for learning with the use of frequent low-stakes testing and the intermixing of various types of testing or retrieval-practice exercises with other types of instructional aids throughout the educational process.

Many courses in both high school and college follow a basic schedule, illustrated on the left side of Fig. 1. Namely, students spend a few weeks, or often the first half of the course, being introduced to topics A–C, followed by a test covering topics A–C, then spend the second half of the course being introduced to topics D–F, and are then tested on topics D–F. Furthermore, these two exams are often heavily weighted (e.g., each exam is worth ~40% of a student's final grade) and often primarily contain only multiple-choice questions. Although this course schedule and format are commonly used and thus familiar to both students and instructors, on the right side of Fig. 1 we illustrate a better way in which tests can be used to enhance students' learning experiences and long-term retention.

When courses contain a small number of tests, with each test accounting for a large portion of a student's course grade, such exams can trigger test anxiety, a form of academic anxiety involving feelings of fear, dread, or nervousness about an upcoming evaluative event (Cassady, 2004, 2010; Wood et al., 2016). Such anxiety

Fig. 1 Example course schedule with two high-stakes exams (a) and frequent testing (b)



can lead to poor academic performance (Cassady & Johnson, 2002; Putwain, 2008; Putwain & Best, 2011; Williams, 1991), but there may be ways to reduce test anxiety while also enhancing learning.

First, given the negative aura that the term testing has now come to evoke among many instructors and students, the use of other terms for this instructional aid—such as low-stakes quizzing, retrieval-practice exercises, or measures of progress—may serve to reduce students' test anxiety (e.g., Agarwal et al., 2014). Additionally, rather than giving only a small number of high-stakes exams, employing many low-stakes exams may reduce students' test anxiety (see Erbe, 2007; Silaj et al., 2021 for work on test anxiety in the classroom). Specifically, such frequent testing can provide numerous opportunities for students to reinforce their knowledge, improving their actual understanding of the material and potentially counteracting feelings of anxiety. Additionally, regular testing allows students to identify and address gaps in their knowledge, which can alleviate anxiety stemming from uncertainty about what they know. Thus, as students observe the benefits of repeated testing, they may view testing as a valuable tool rather than something to stress over.

Repeated testing can also harness the benefits of the testing effect to maximize learning. For example, prior work has demonstrated that more frequent exams are associated with better learning outcomes (e.g., Bangert-Drowns et al., 1991; Leeming, 2002; see also Roediger & Karpicke, 2006a). More specifically, when Leeming (2002) compared students who took a short exam at the beginning of every class with students in classes that had only a few exams for the same material, students in the exam-a-day classes achieved significantly better grades, were less likely to drop the class, and performed better on a later test. Furthermore, anonymous questionnaires revealed that most students believed that having an exam every day led to their doing more studying and achieving better learning as compared to their other classes (and students also reported liking this procedure). Thus, frequent exams—and

especially ones that not only ask questions about the just presented block of material but also include a few questions from previous blocks, as illustrated in Fig. 1b—may positively impact student performance, retention, and perceptions of their learning.

The use of different forms of low-stakes testing, such as polling questions (e.g., multiple-choice questions presented electronically via applications like Poll Everywhere, Mentimeter, or responded to with electronic iClicker remotes) or review games (e.g., using applications like Kahoot! or Google Forms), can benefit learners in multiple ways (e.g., Deslauriers et al., 2011; Pan et al., 2019). Firstly, it promotes active engagement, retrieval practice, and feedback, as many forms of low-stakes testing provide immediate feedback to learners, helping them identify and correct misconceptions or errors. Additionally, as previously mentioned, low-stakes testing seems to reduce test anxiety, creating a relaxed and positive learning environment where learners feel more comfortable taking risks, making mistakes, and learning from them. Moreover, low-stakes testing may increase learners' motivation to study and prepare for assessments as it provides opportunities for them to see the immediate results of their efforts, leading to a sense of achievement and satisfaction.

Employing frequent tests can also capture the benefits of the spacing effect: when study time is distributed rather than massed, long-term memory is improved (Bjork & Allen, 1970; Cepeda et al., 2006; Greene, 2008; Karpicke & Bauernschmidt, 2011; Murphy et al., 2022; see Carpenter, 2017 for a review). Specifically, we can induce our students to space their studying and learning activities by using more frequent tests as opposed to having them resort to cramming before high-stakes exams (Fitch et al., 1951), which may support short-term performance but does not lead to long-term learning. Additionally, more frequent tests may result in the same information being tested twice (assuming exams are cumulative to some degree, as represented in Fig. 1b), which should result in accruing the benefits of spaced retrieval (Balota et al., 2007). As such, although cumulative exams are often disliked by students, cumulative exams can be more beneficial for their learning than non-cumulative exams (Lawrence, 2013) by harnessing both the testing effect (i.e., frequent testing of earlier course material) and the spacing effect (i.e., students' revisiting previously learned concepts during their preparation for cumulative exams). Thus, incorporating frequent tests that are cumulative, at least to some extent, can leverage both the benefits of retrieval practice and spacing.

Just as we advocate the use of tests as providing beneficial retrieval practice, we also believe that a balanced and thoughtful grading approach (how to weigh each course activity as it relates to students' grades) is essential. By considering the current research on grading and retrieval practice in real-world educational contexts, instructors can make informed decisions to create supportive learning environments that maximize student learning and minimize test anxiety. We encourage further investigation into grading approaches and their impact on learning outcomes so instructors can implement evidence-based practices that promote meaningful and lasting learning.

What Kind of Test Formats Are Best?

Although we have so far extolled the benefits of testing or retrieval practice for enhancing learning, instructors need to be aware that not all types of tests or retrieval practice exercises produce the same benefits for learning. For example, while the multiple-choice format is more practical to use in large classes due to the ease and efficiency with which such questions can be graded (thereby lessening the time before feedback can be provided to students), instructors need to create such questions in a way that they require active retrieval on the part of the students.

To do so, multiple-choice questions need to provide the student with a set of competitive alternatives (i.e., alternatives that are plausible enough to be possible correct answers) so that students need to retrieve information about each alternative to select the correct one as opposed to being able to easily recognize a correct answer from, say, a set of alternatives that are mostly non-competitive or implausible possibilities. In other words, to produce enhanced learning, instructors need to create the type of multiple-choice questions that require students to engage in active retrieval processes. For example, imagine a question about the name of the Greek goddess of love (*answer*: Aphrodite). The names of other Greek and Roman goddesses (e.g., Venus, Hera, and Athena) would be more competitive than the names of Greek and Roman gods (e.g., Zeus, Mars, and Hades) or names that are not even Greek or Roman gods or goddesses. Here, the names of other Greek and Roman goddesses are more plausible as the correct answer and students may need to think about why such alternatives are wrong (e.g., Venus is the Roman goddess of love) to reject them (see Little et al., 2019).

It is important to note that while all competitive alternatives are plausible, not all plausible alternatives are necessarily competitive. In the context of multiple-choice questions with competitive alternatives, competitive alternatives are those that require students to retrieve information about each option to determine the correct answer. This process of active retrieval enhances learning and can lead to better performance on both previously asked questions and related questions. Plausible alternatives, on the other hand, simply answer choices that make sense in the context of the question and could be seen as potentially correct, but they may not require the same level of retrieval as competitive alternatives. To develop competitive alternatives, instructors need to ensure that each alternative is based on information that is closely related to the correct answer, thus requiring students to engage in retrieval processes. On the contrary, plausible alternatives may not be related in such a way that prompts active retrieval. However, it is important to strike a balance between providing competitive alternatives that challenge students without making the questions overly difficult or confusing.

Competitive multiple-choice questions can also enhance students' ability to answer questions about one of the formerly incorrect alternatives on a later exam (Little & Bjork, 2015; Little et al., 2012). That is, such multiple-choice questions can enhance later performance for both previously asked questions and new related questions. This advantage is thought to arise because when competitive

alternatives are provided, students try to retrieve what they have learned about each alternative, and this effort then not only strengthens what they have previously heard or read about the correct choice but also strengthens what they have previously heard or read about each of the competitive alternatives (Little & Bjork, 2015; Little et al., 2019).

To test this possible explanation, Little and Bjork (2015) had students read lessons on the solar system and ferrets before completing a practice multiple-choice test for one of those topics. On the test, half of the questions had competitive alternatives and half had non-competitive alternatives. For example, some participants might answer, “What is the hottest terrestrial planet?” with the choices Venus, Mars, and Mercury (competitive alternatives), while other participants were required to answer that same question but with Venus, Uranus, and Saturn as choices (non-competitive alternatives in that neither Uranus nor Saturn are terrestrial planets). Additionally, if the Venus question had appeared as a competitive question, participants would have also received a question about Neptune that was competitive, with Saturn and Uranus as choices, and if the Venus question had been presented as a non-competitive question, participants would have received a question about Neptune with Mars and Mercury as choices. On a later delayed exam, students were significantly better at answering new questions about the alternatives (e.g., Which planet was first visited by Mariner 10? *Answer*: Mercury; Which planet’s axial tilt is 90° to the plane of its orbit? *Answer*: Uranus) when those alternatives had been included as competitive alternatives than when they had not been.

Follow-up research used a procedure in which participants were asked to report what they were thinking when they answered such multiple-choice questions (Little et al., 2019). Most participants reported at least occasionally using an elimination strategy, and in some cases, participants spontaneously reported recalling information about the incorrect alternatives to reject them. When participants recalled information about the incorrect alternative and then that alternative was the correct answer to a question appearing on a later cued-recall test, such participants were very likely to correctly answer that question. Thus, the implementation of appropriate incorrect alternatives for multiple-choice questions is an important component of writing questions that can produce enhanced learning for both information that is directly tested and information that is related to that question’s correct answer but is not directly tested.

Besides competitive multiple-choice questions, other forms of questions can enhance learning. For example, questions requiring the student to engage in generation processes as part of obtaining the correct answer can benefit learning. Specifically, students’ later performance will be enhanced because it will benefit from the generation effect: better long-term memory when learners take an active part in producing the information they are to learn. Applied to assessment, instructors should incorporate more opportunities for students to generate the to-be-learned material (e.g., short answer questions, fill-in-the-blank questions, etc.; examples of such learning tasks appear in DeWinstanley & Bjork, 2004; Hertel, 1989).

Cued-recall, short-answer, and fill-in-the-blank types of questions are prime examples of the types of test questions that require active retrieval processes on the part of students and, thus, can serve as tools for learning as well as assessment.

Questions employing this format tend to be relatively easy for instructors to write and have traditionally been considered more favorably by educators than those employing a multiple-choice format. However, short-answer questions can take significantly more time to grade than most instructors have available. Fortunately, several studies conducted in the laboratory have shown that using competitive multiple-choice questions, where all the answer choices are plausible options, can be just as effective in improving students' performance on subsequent cued-recall exams as practice tests using cued-recall or short-answer questions (Little et al., 2012). Furthermore, McDaniel and Little (2019) have suggested that competitive multiple-choice and short-answer quizzing can be equally effective in the classroom.

In sum, both short-answer questions and well-designed multiple-choice questions can serve as effective tools for enhancing learning. There is one consideration, however, that might indicate that the use of well-designed multiple-choice questions would be better for enhancing students' learning than the use of short or cued-recall questions. In contrast to multiple-choice questions with competitive alternatives, short answer or cued-recall tests tend to focus attention only on the question at hand—possibly prompting individuals to try to ignore competing information—thus setting up conditions for the possibility of retrieval-induced forgetting.

Retrieval-induced forgetting refers to the finding that cued-recall tests, where students are given cues to recall information from memory, can sometimes impair their ability to later answer questions involving related information (Anderson et al., 1994). Although most often shown with cued-recall pairs, this effect has also sometimes been shown with educational materials (Chan, 2009; Little et al., 2011, 2012). Thus, while cued-recall practice tests can be effective in enhancing memory for the practiced items, they may also lead to the inhibition or suppression of competitive, related, but non-practiced information, resulting in retrieval-induced forgetting of that information¹. In other words, trying to recall specific information during a cued-recall practice test can unintentionally impair memory for competitive, related information, which can hinder students' ability to answer questions about that related information in subsequent tests or assessments. Such results highlight the complex and sometimes counterintuitive nature of memory processes and the need for careful consideration of the types of practice tests used in educational settings.

Although including short-answer questions or more competitive multiple-choice tests in our instructional practices can be beneficial for our students' learning, short-answer questions can be difficult and time-consuming to grade, and creating competitive multiple-choice tests can be difficult and time-consuming to create, particularly as compared to their non-competitive counterparts. Thus, even instructors who are eager to use short-answer or competitive

¹ Note that retrieval-induced forgetting with educational materials depends upon various factors, one of which is a competitive relationship between the tested and related content. When the tested and related content is not competitive, cued-recall testing on some information can improve recall of other information (see Anderson & Biddle, 1975; Hamaker, 1986, for a review).

multiple-choice tests are sometimes thwarted in their efforts to do so simply because of the difficulty in grading short-answer questions or in coming up with four or five competitive alternatives to include in each competitive multiple-choice question. Fortunately, recent work has demonstrated that true-false questions can have some of the same beneficial effects as competitive multiple-choice questions (Brabec et al., 2021).

Competitive true-false questions can produce better later performance for both previously asked questions and related questions. For example, suppose students have just had a lesson on Yellowstone Park that included a discussion of how geysers work and some of the famous geysers to be found there. A simple example of a competitive true-false question would be “True or False: Steamboat Geyser, not Castle Geyser, is the oldest geyser in Yellowstone Park.” To answer this question (which is false), students appear to retrieve both what they have learned about Steamboat Geyser and what they have learned about Castle Geyser, resulting in a better ability to answer questions about either one of these geysers on a later exam. Thus, true-false questions of this type, which are much easier to write, may offer similar benefits to multiple-choice questions with competitive alternatives.

In sum, multiple-choice questions with competitive alternatives, despite often being challenging and time-consuming to write, can improve learning outcomes by prompting students to recall information about all the alternatives, leading to retrieval practice benefits when answering later questions concerning any of the alternatives. However, if instructors do not have the time required to write competitive multiple-choice questions, competitive true-false questions can provide a solution—they too can increase the students’ learning of or access to the correct answers for both previously asked and related questions. Such findings indicate that when properly constructed, multiple-choice and true/false questions can both be powerful tools for promoting learning, challenging the notion that multiple-choice or true/false questions are inferior to cued-recall questions.

Should Students Take Tests Independently?

Some research has examined the benefits of group versus individual testing. For example, Cranney et al. (2009) had first-year college students watch a psychobiology video followed by a video-related activity and then a surprise test that they took individually. Looking at performance on the surprise test, the researchers compared the effectiveness of a group quiz, an individual quiz, a restudy condition, and a no-activity control condition. In general, results indicated that taking quizzes yielded better outcomes than not taking quizzes, and interestingly, the group quiz condition outperformed the individual quiz condition.

Collaborative testing can take various forms, and one such strategy involves the individual taking a first quiz, which is then followed by the opportunity to complete the same quiz in small groups, with the group performance contributing to some portion of the student’s grade (e.g., Rao et al., 2002). Using this type of procedure (i.e., an individual test followed by either an individual retest or a group retest), Gilley

and Clarkston (2014) showed that the taking of a group retest was more effective for learning (as evaluated through a later individual test) than the individual taking the retest. Moreover, students generally enjoy collaborative testing and report reduced test anxiety (e.g., Lusk & Conklin, 2003). However, research on group testing versus individual testing has yielded mixed results, with some studies showing that group testing is not superior to individual testing for long-term retention and transfer (e.g., LoGiudice et al., 2015; Lusk & Conklin, 2003; Vojdanoska et al., 2010; Wissman & Rawson, 2018).

In certain conditions, group testing might even be worse, which aligns with the concept of collaborative inhibition, which occurs when groups of individuals collectively recall and remember information less accurately compared to if they had worked alone. To use collaborative testing in an educational context, it is essential to consider that collaborative inhibition is more likely to occur with open-ended retrieval, whereas tests with more specific cues like cued-recall or multiple-choice (which are common in educational contexts and especially in the review activities discussed in this commentary) are less likely to lead to collaborative inhibition (see Rajaram & Pereira-Pasarin, 2010 for a review of conditions promoting collaborative inhibition vs. facilitation; see also LoGiudice et al., 2015 for an educational review on collaborative testing). Taking all these findings into account, collaborative testing in the educational settings we have discussed may be advantageous and, at worst, is unlikely to be detrimental. Furthermore, it is also a procedure that appeals to students. Thus, incorporating collaborative retrieval activities, such as interactive games and test-taking, into one's instructional teaching strategies can be a motivating way to engage students in practices that should facilitate their learning.

When Should We Give Tests?

Testing not only assesses what students know but also enhances their ability to learn new material in subsequent study sessions. Specifically, if students are asked to answer questions about a passage they are about to read or a lesson they are about to be given, their learning of the then-presented material is enhanced even if they are not able to answer any of those questions correctly (e.g., Arnold & McDermott, 2013; Hays et al., 2013; Richland et al., 2009). Thus, instructors should consider administering pre-tests prior to instruction to enhance long-term learning.

The extent of this pretesting advantage (see Bjork et al., 2015; Carpenter & Toftness, 2017; Carpenter et al., 2018, 2023; Sana & Carpenter, 2023) can depend on the type of testing format used in the pretests. For example, using both multiple-choice and cued-recall test formats, Little and Bjork (2016) examined the effects of using tests as pretests (i.e., before studying) on the subsequent learning of information related to the correct answers on the pretest but not the specific correct answer itself. Overall, results revealed that multiple-choice pretesting was more effective than cued-recall pretesting, even after a delay. Specifically, both test types enhanced the learning of the tested content, but multiple-choice pretesting also enhanced the

learning of the subsequently presented related information more so than did cued-recall pretesting. This may be because multiple-choice tests made students pay attention to both the correct answer and other related details when they came across them again (see Carpenter et al., 2023 for a review of the benefits of prequestions/pretests).

While the nature of the processes underlying the benefits of pretesting is still being debated, it is fairly widely agreed that a major reason for this benefit is that pretesting leads students to think more deeply and critically about the information that was pretested when it is later encountered during the presentation of the to-be-learned material, resulting in a more elaborate encoding of such material. For example, even for questions to which students do not already know the correct answer, if they are required to search their memories for possible answers to such questions before being allowed to search for them on the Internet, they will remember the found answers better than if they had been allowed to search for them immediately (Giebl et al., 2021, 2022). Additionally, pretests can lead to a reduction in mind wandering (Pan et al., 2020) and enhance students' capacity to maintain focus during lessons (Pan & Sana, 2021). Thus, instructors should consider giving tests before lessons as another method of using tests as a means for potentiating their students' learning.

To summarize, considerable evidence suggests that pretests can enhance learning when they require students to attempt retrieval, even if the correct answer is not successfully recalled. As a result, we recommend the use of pretests given before the presentation of the to-be-learned material using either multiple-choice questions with competitive alternatives or competitive true-false questions, both of which have been shown to benefit subsequent learning outcomes for both the tested and related information.

What Issues Require More Research?

Although the effects of testing in the reviewed literature are robust, we need to do more to examine the generalizability of these benefits. For example, a recent review of 50 classroom experiments by Agarwal et al. (2021) demonstrated that retrieval practice yields medium to large benefits in most cases (57%), and the positive impact of retrieval practice on learning was observed across various education levels, content areas, experimental designs, final test delays, retrieval and final test formats, and timing of retrieval practice and feedback. However, the review also highlights that only a small fraction of experiments (6%) were conducted in non-Western, educated, industrialized, rich, and democratic (non-WEIRD) countries. Thus, while retrieval practice has been shown to offer substantial benefits for learning across many educational settings, whether such benefits accrue across even more diverse educational contexts remains to be determined. Additionally, more specific research needs to be conducted regarding how individual differences such as students' prior knowledge, cultural backgrounds, and socioeconomic status influence how retrieval practice impacts learning. The results of such investigations should provide instructors with additional information regarding how

testing might be used to foster more equitable educational experiences and outcomes for all students.

Testing and other forms of active learning have been consistently shown to benefit students of all abilities and can be particularly advantageous for capable but underperforming students (Haak et al., 2011). For example, a review conducted by Theobald et al. (2020) analyzed studies comparing the performance of underrepresented students (e.g., low-income, ethnic minority, or racial minority) to their overrepresented peers in both active learning and traditional instructional settings. Results revealed that active learning approaches tended to narrow the achievement gaps between these groups. Thus, incorporating question-answering activities as a form of active learning into one's instructional practices would seem to hold the potential to be one way to promote greater equity in education and reduce achievement gaps among different student populations.

While the current commentary emphasizes the benefits of testing for learning and takes the position that these benefits may serve as a potential "equalizer" in enhancing learning outcomes for all students, there is a need for further investigation to understand the implications of testing in different academic disciplines, particularly in the context of addressing equity gaps. The existing research on equity gaps has predominantly focused on Science, Technology, Engineering, and Mathematics (STEM) disciplines, where the underrepresentation of certain groups, particularly women and minorities, remains a concern. It is essential to explore more thoroughly how testing might contribute to reducing these disparities and whether any such contributions might vary across different subject areas.

One critical aspect of future research should involve comparing the effectiveness of testing in both STEM and non-STEM disciplines. While the benefits of active learning and testing have been demonstrated across various subjects, it is essential to understand if the potential role of testing as an "equalizer" differs between these disciplines. Investigating the impact of testing on students' academic achievement and retention rates in non-STEM fields will provide valuable insights into its broader applicability and potential to enhance learning outcomes more universally.

Furthermore, research exploring the combination of testing with other active learning strategies in different academic domains should be undertaken. While this commentary primarily focuses on the use of testing to involve students in active learning, it is important to acknowledge that active learning encompasses a range of instructional approaches. Future studies could examine how the integration of testing with other interactive activities influences student engagement, motivation, and learning in STEM and non-STEM disciplines. The discovery of potential synergistic effects when different active learning strategies are combined may lead to the development of more effective and comprehensive instructional practices.

How Can We Implement these Principles in the Classroom?

Again, while the administration of tests is already very common in the classroom as a means of assessing learning, we argue in this commentary that instructors should also be using tests to potentiate the learning of their students, and we have

Table 1 Recommendations for testing in the classroom

-
- 1) Test frequently rather than infrequently, and in addition to exams, use polling questions, review games, and quizzes in addition to exams
 - 2) Use tests that require retrieval processes. For multiple-choice questions to require retrieval, the incorrect alternatives should be plausible and competitive
 - 3) Provide feedback
 - 4) Incorporate collaborative retrieval
 - 5) Consider pretests in addition to post-tests
 - 6) Encourage students to self-test
-

summarized the various ways in which doing so can be accomplished in Table 1. For example, rather than the only tests in a course being a midterm and final exam—as illustrated on the left side of Fig. 1, which represents a common organization of many courses—instructors should include many low-stakes exams whose main purpose is to enhance learning (as illustrated in the right side of Fig. 1b). In short, the more tests we give our students on the information we are trying to teach them, whether given before or after learning, the more likely our students will be to remember that information later and be able to use it in different contexts.

In the classroom, an instructor has the option to employ various testing tools such as clickers or polling questions, review games (to be completed individually, collaboratively, or in a combination of both), and quizzes. For instance, one of the authors of this paper utilizes Google Forms to create collaborative quizzes for students. A notable advantage of using Google Forms is the instant availability of quiz answers to the instructor and instantaneous graphs of results that are easy to show students, allowing for immediate performance observation and feedback provision. They are also easy to use both in class and during online teaching sessions. This real-time feedback could enhance the learning experience for students and aid instructors in gauging students' progress effectively.

In addition to introducing more desirable difficulties, such as tests or retrieval practice, into our instructional efforts, we also need to teach our students how to introduce desirable difficulties into their own study practices. With respect to their profiting from the testing effect, we should encourage our students to engage in self-testing as much as possible. Doing so can take the form of asking students to write down the main points from a chapter they have just read without looking back at it, summarizing the main points from a lecture right after class without looking at any notes, or getting together in small study groups where the students practice testing one another—an activity that many students already report doing (Wissman & Rawson, 2016). Students should also be encouraged to use any testing resources provided by their textbook. The more students engage in activities that test their learning or require them to generate aspects of the to-be-learned material, the more likely they are to begin to appreciate the benefits of testing (as well as other desirable difficulties) for enhancing their learning, even though engaging in desirable difficulties can require more effort on the part of the learner.

How Can We Overcome Barriers to Implementation?

Despite the numerous lab- and classroom-based studies demonstrating the benefits of desirable difficulties like the testing effect (see Rowland, 2014; Schwieren et al., 2017 for reviews), many obstacles are encountered when trying to introduce desirable difficulties into various types of educational settings—even when both instructors and students want to do so (see Bjork & Bjork, 2022 for a discussion of these obstacles). As the name indicates, desirable difficulties present difficulties or challenges for learners (e.g., it is much easier simply to restudy information than to test yourself on it) and they can often slow down the rate at which one's performance improves, which can be mistakenly interpreted by students (and instructors as well) as impairing the learning process. Moreover, some desirable difficulties defy conventional wisdom and can seem at odds with the types of teaching or instruction with which both students and instructors have become familiar. Lastly, students may not want to change their approach to the learning process if they have had prior academic success (i.e., they have been able to earn good grades) without using desirable difficulties.

Instructors may have reservations about incorporating more testing into their teaching for reasons other than those just discussed. Two main additional reasons seem to be: (a) they fear it takes away valuable time that could be used for content delivery or restudying; and (b) they worry about the increased workload involved in implementing testing, such as writing more exams, incorporating polling questions, and grading. However, research has consistently demonstrated that testing actually enhances learning more than control conditions that match time on task (e.g., Roediger & Karpicke, 2006b). In other words, the time invested in testing is not wasted but rather contributes significantly to improved learning outcomes.

To facilitate the implementation of testing and other effective strategies, we recommend the use of available resources and technologies that can streamline the process. For instance, employing digital tools like quiz generators or learning management systems as well as the test banks provided with many textbooks can significantly reduce the burden of test preparation and grading, allowing instructors to focus on other aspects of their teaching. Additionally, providing instructors with clear guidelines, sample questions, and templates for creating tests can expedite the process and make it more manageable.

Despite the potential increase in the instructors' workload, we believe that the benefits to our students make the effort worthwhile. The incorporation of more testing and interactive elements in teaching fosters active learning and enhances students' retention and comprehension of the material. While instructors may feel the need to update higher-stakes assessments each semester to maintain their integrity and avoid potential cheating, the same level of urgency may not be necessary for lower-stakes assessments like polling and review games. Once these assessment questions are integrated into the lecture materials, instructors may find that they require minimal additional work from semester to semester. As a result, the time and effort invested in creating these interactive assessments can prove to be a valuable and sustainable resource in the long run, benefiting both instructors and students alike. Ultimately, the positive impact on students' academic performance and long-term learning justifies the additional effort required by instructors.

Conclusions

As we try to educate both more students and a broader range of students than we have traditionally done in the past, we believe it is essential for instructors to give students the knowledge and ability to incorporate desirable difficulties into their study strategies and their self-guided learning activities. Among other reasons, there is growing evidence that tasks involving active learning—of which we believe testing is one—can serve as an equalizer for our students (e.g., Haak et al., 2011; Theobald et al., 2020). That is, regardless of the many individual differences among students and the great variance in the level of preparation students may have at the start of any educational endeavor, the knowledge of how to use desirable difficulties to improve their study strategies can enable all students to succeed. We hope that the present commentary can help make both students and instructors more aware of the benefits of testing for achieving learning that is both long-lasting and transferable, which is the ultimate goal of education.

Declarations

Conflict of Interest The authors declare no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Agarwal, P. K., D'Antonio, L., Roediger, H. L., McDermott, K. B., & McDaniel, M. A. (2014). Classroom-based programs of retrieval practice reduce middle school and high school students' test anxiety. *Journal of Applied Research in Memory and Cognition*, 3, 131–139.
- Agarwal, P. K., Nunes, L. D., & Blunt, J. R. (2021). Retrieval practice consistently benefits student learning: A systematic review of applied research in schools and classrooms. *Educational Psychology Review*, 33, 1409–1453.
- Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1063–1087.
- Anderson, R. C., & Biddle, W. B. (1975). On asking people questions about what they are reading. In *Psychology of Learning and Motivation* (Vol. 9, pp. 89–132). Academic Press.
- Arnold, K. M., & McDermott, K. B. (2013). Test-potentiated learning: Distinguishing between direct and indirect effects of tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39, 940–945.
- Balota, D. A., Duchek, J. M., & Logan, J. M. (2007). Is expanded retrieval practice a superior form of spaced retrieval? A critical review of the extant literature. In J. S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger* (pp. 83–105). Psychology Press.

- Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C. L. C. (1991). Effects of frequent classroom testing. *Journal of Educational Research, 85*, 89–99.
- Bjork, E. L., & Bjork, R. A. (2014). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher & J. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (2nd ed., pp. 59–68).
- Bjork, E. L., & Bjork, R. A. (2022). Introducing desirable difficulties into practice and instruction: Obstacles and opportunities. In C. E. Overson, C. M., Hakala, L. L. Kordonowy, and V. A. Benassi (Eds.), *What scholars and teachers want you to know about why and how to apply the science of learning in your academic setting*.
- Bjork, E. L., Soderstrom, N. C., & Little, J. L. (2015). Can multiple-choice testing induce desirable difficulties? Evidence from the laboratory and the classroom. *The American Journal of Psychology, 128*, 229–239.
- Bjork, R. A. (1975). Retrieval as a memory modifier. In R. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 123–144). Lawrence Erlbaum Associates.
- Bjork, R. A., & Allen, T. W. (1970). The spacing effect: Consolidation or differential encoding? *Journal of Verbal Learning and Verbal Behavior, 9*, 567–572.
- Brabec, J. A., Pan, S. C., Bjork, E. L., & Bjork, R. A. (2021). True-false testing on trial: Guilty as charged or falsely accused? *Educational Psychology Review, 33*, 667–692.
- Carpenter, S. K. (2017). Spacing effects in learning and memory. In J. T. Wixted (Ed.), *Cognitive Psychology of Memory, Vol. 2 Learning and Memory: A Comprehensive Reference, 2nd edition*, J. H. Byrne (Ed.), (pp. 465–485). Academic Press.
- Carpenter, S. K., King-Shepard, Q., & Nokes-Malach, T. J. (2023). The prequestion effect: Why it is useful to ask students questions before they learn. In C. Overson, C. Hakala, L. Kordonowy, & V. Benassi (Eds.), *In their own words: What scholars want you to know about why and how to apply the science of learning in your academic setting* (pp. 74–82). Society for the Teaching of Psychology.
- Carpenter, S. K., Rahman, S., & Perkins, K. (2018). The effects of prequestions on classroom learning. *Journal of Experimental Psychology: Applied, 24*, 34–42.
- 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*, 104–109.
- Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition, 20*, 633–642.
- Cassady, J. C. (2004). The impact of test anxiety on text comprehension and recall in the absence of external evaluative pressure. *Applied Cognitive Psychology, 18*, 311–325.
- Cassady, J. C. (2010). Test anxiety: Contemporary theories and implications for learning. In J. C. Cassady (Ed.), *Anxiety in schools: The causes, consequences, and solutions for academic anxieties* (pp. 7–26). Peter Lang.
- Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Anxiety, 27*, 270–295.
- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*, 354–380.
- Chan, J. C. (2009). When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing. *Journal of Memory and Language, 61*, 153–170.
- Cranney, J., Ahn, M., McKinnon, R., Morris, S., & Watts, K. (2009). The testing effect, collaborative learning, and retrieval-induced facilitation in a classroom setting. *European Journal of Cognitive Psychology, 21*, 919–940.
- Deslauriers, L., Schelew, E., & Wieman, C. (2011). Improved learning in a large-enrollment physics class. *Science, 332*, 862–864.
- DeWinstanley, P. A., & Bjork, E. L. (2004). Processing strategies and the generation effect: Implications for making a better reader. *Memory & Cognition, 32*, 945–955.
- Dunlosky, J., & Hertzog, C. (1998). Training programs to improve learning in later adulthood: Helping older adults educate themselves. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 249–275). Erlbaum.
- Erbe, B. (2007). Reducing test anxiety while increasing learning: The cheat sheet. *College Teaching, 55*, 96–98.
- Fitch, M. L., Drucker, A. J., & Norton, J. A. (1951). Frequent testing as a motivating factor in large lecture courses. *Journal of Educational Psychology, 42*, 1–20.
- Giebl, S., Mena, S., Sandberg, R., Bjork, E. L., & Bjork, R. A. (2022). Thinking first versus googling first: Preferences and consequences. *Journal of Applied Research in Memory and Cognition*.

- Giebl, S., Mena, S., Storm, B. C., Bjork, E. L., & Bjork, R. A. (2021). Answer first or Google first? Using the Internet in ways that enhance, not impair, one's subsequent retention of needed information. *Psychology Learning & Teaching, 20*, 58–75.
- Gilley, B. H., & Clarkston, B. (2014). Collaborative testing: Evidence of learning in a controlled in-class study of undergraduate students. *Journal of College Science Teaching, 43*, 83–91.
- Greene, R. L. (2008). Repetition and spacing effects. In H. L. Roediger (Ed.), *Learning and memory: A comprehensive reference. Cognitive psychology of memory* (Vol. 2, pp. 65–78). Elsevier.
- Haak, D. C., HilleRisLambers, J., Pitre, E., & Freeman, S. (2011). Increased structure and active learning reduce the achievement gap in introductory biology. *Science, 332*, 1213–1216.
- Halamish, V., & Bjork, R. A. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 801–812.
- Hamaker, C. (1986). The effects of adjunct questions on prose learning. *Review of Educational Research, 56*, 212–242.
- Hays, M. J., Kornell, N., & Bjork, R. A. (2013). When and why a failed test potentiates the effectiveness of subsequent study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 290–296.
- Hertel, P. T. (1989). The generation effect: A reflection of cognitive effort? *Bulletin of the Psychonomic Society, 27*, 541–544.
- Imundo, M. N., Pan, S. C., Bjork, E. L., & Bjork, R. A. (2021). Where and how to learn: The interactive benefits of contextual variation, restudying, and retrieval practice for learning. *Quarterly Journal of Experimental Psychology, 74*, 413–424.
- Karpicke, J. D. (2017). Retrieval-based learning: A decade of progress. In J. T. Wixted (Ed.), *Cognitive psychology of memory, of learning and memory: A comprehensive reference* (Vol. 2, pp. 487–514). Academic Press.
- Karpicke, J. D., & Bauernschmidt, A. (2011). Spaced retrieval: Absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 1250–1257.
- Kornell, N., & Bjork, R. A. (2008). Learning Concepts and Categories: Is Spacing the “Enemy of Induction”? *Psychological Science, 19*, 585–592.
- Lawrence, N. K. (2013). Cumulative exams in the introductory psychology course. *Teaching of Psychology, 40*, 15–19.
- Leeming, F. C. (2002). The exam-a-day procedure improves performance in psychology classes. *Teaching of Psychology, 29*, 210–212.
- Little, J. L., & Bjork, E. L. (2015). Optimizing multiple-choice tests as tools for learning. *Memory & Cognition, 43*, 14–26.
- Little, J. L., & Bjork, E. L. (2016). Multiple-choice pretesting potentiates learning of related information. *Memory & Cognition, 44*, 1085–1101.
- Little, J. L., Bjork, E. L., Bjork, R. A., & Angello, G. (2012). Multiple-choice tests exonerated, at least of some charges: Fostering test-induced learning and avoiding test-induced forgetting. *Psychological Science, 23*, 1337–1344.
- Little, J. L., Frickey, E. A., & Fung, A. K. (2019). The role of retrieval in answering multiple-choice questions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*, 1473–1485.
- Little, J. L., & McDaniel, M. A. (2015). Metamemory monitoring and control following retrieval practice for text. *Memory & Cognition, 43*, 85–98.
- Little, J. L., Storm, B. C., & Bjork, E. L. (2011). The costs and benefits of testing text materials. *Memory, 19*, 346–359.
- LoGiudice, A. B., Pachai, A. A., & Kim, J. A. (2015). Testing together: When do students learn more through collaborative tests? *Scholarship of Teaching and Learning in Psychology, 1*, 377–389.
- Lusk, M., & Conklin, L. (2003). Collaborative testing to promote learning. *Journal of Nursing Education, 42*, 121–124.
- McDaniel, M. A., & Little, J. L. (2019). Multiple-choice and short-answer quizzing on equal footing in the classroom: Potential indirect effects of testing. In J. Dunlosky & K. A. Rawson (Eds.), *The Cambridge handbook of cognition and education* (pp. 480–499). Cambridge University Press.
- McDaniel, M. A., & Masson, M. E. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 371–385.
- Murphy, D. H., Bjork, R. A., & Bjork, E. L. (2022). Going beyond the spacing effect: Does it matter how time on a task is distributed? *Quarterly Journal of Experimental Psychology*.

- Narens, L., Nelson, T. O., & Shecke, P. (2008). Memory monitoring and the delayed JOL effect. *Handbook of metamemory and memory*, 137–153.
- Pan, S. C., Cooke, J., Little, J. L., McDaniel, M. A., Foster, E. R., Connor, L. T., & Rickard, T. C. (2019). Online and clicker quizzing on jargon terms enhances definition-focused but not conceptually focused biology exam performance. *CBE—Life Sciences Education*, 18, ar54.
- Pan, S. C., & Sana, F. (2021). Pretesting versus posttesting: Comparing the pedagogical benefits of errorful generation and retrieval practice. *Journal of Experimental Psychology: Applied*, 27, 237–257.
- Pan, S. C., Sana, F., Schmitt, A., & Bjork, E. L. (2020). Pretesting reduces mind wandering and enhances learning during online lectures. *Journal of Applied Research in Memory and Cognition*, 9, 542–554.
- Putwain, D. W. (2008). Deconstructing test anxiety. *Emotional and Behavioural Difficulties*, 13, 141–155.
- Putwain, D. W., & Best, N. (2011). Fear appeals in the primary classroom: Effects on test anxiety and test grade. *Learning and Individual Differences*, 21, 580–584.
- Rajaram, S., & Pereira-Pasarin, L. P. (2010). Collaborative memory: Cognitive research and theory. *Perspectives on Psychological Science*, 5, 649–663.
- Rao, S. P., Collins, H. L., & DiCarlo, S. E. (2002). Collaborative testing enhances student learning. *Advances in Physiology Education*, 26, 37–41.
- Rhodes, M. G. (2016). Judgments of learning. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 65–80). Oxford University Press.
- Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied*, 15, 243–257.
- Roediger, H. L., & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1, 181–210.
- Roediger, H. L., & Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17, 249.
- Roediger, H. L., Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. *Psychology of Learning and Motivation*, 55, 1–36.
- Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science*, 35, 481–498.
- Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, 140, 1432–1463.
- Sana, F., & Carpenter, S. K. (2023). Broader benefits of the pretesting effect: Placement matters. *Psychonomic Bulletin & Review*.
- Schwieren, J., Barenberg, J., & Dutke, S. (2017). The testing effect in the psychology classroom: A meta-analytic perspective. *Psychology Learning & Teaching*, 16, 179–196.
- Silaj, K. M., Schwartz, S. T., Siegel, A. L., & Castel, A. D. (2021). Test anxiety and metacognitive performance in the classroom. *Educational Psychology Review*, 33, 1809–1834.
- Smith, S. M., Glenberg, A. M., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, 6, 342–353.
- Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., ... & Freeman, S. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences*, 117, 6476–6483.
- Vojdanoska, M., Cranney, J., & Newell, B. R. (2010). The testing effect: The role of feedback and collaboration in a tertiary classroom setting. *Applied Cognitive Psychology*, 24, 1183–1195.
- Williams, J. E. (1991). Modeling test anxiety, self-concept and high school students' academic achievement. *Journal of Research & Development in Education*, 25, 51–57.
- Wissman, K. T., & Rawson, K. A. (2016). How do students implement collaborative testing in real-world contexts? *Memory*, 24, 223–239.
- Wissman, K. T., & Rawson, K. A. (2018). Collaborative testing for key-term definitions under representative conditions: Efficiency costs and no learning benefits. *Memory & Cognition*, 46, 148–157.
- Wood, S. G., Hart, S., Little, S., & Phillips, S. A. (2016). Test anxiety and a high-stakes standardized reading comprehension test: A behavioral genetics perspective. *Merrill-Palmer Quarterly*, 62, 233–251.