Test-Enhanced Learning

David R. Shanks, Division of Psychology and Language Sciences, University College London, Hilary J. Don, University College London, Shaun Boustani, University College London, and Chunliang Yang, Beijing Normal University

https://doi.org/10.1093/acrefore/9780190236557.013.908
Published online: 22 March 2023

Summary

Tests following learning serve several important functions, including enabling students to monitor their progress and identify knowledge gaps, but they are also learning events in their own right. Testing is a powerful strategy to consolidate retention of studied information, by comparison with restudying and other elaborative strategies, and facilitates subsequent learning of new information. Moreover, the testing effect generalizes to different test formats, study-test intervals, and material types, and has been robustly demonstrated not only in the laboratory but also in classroom settings. Pretesting can promote subsequent learning of tested information, but its effect on non-pre-questioned information remains unclear. Although the beneficial effects of testing on learning and memory are substantial, learners tend to underappreciate the merits of practice tests, leading to their underemployment. Lack of motivation or insufficient knowledge about how best to exploit testing may be factors that suppress its use. However, some promising interventions have been developed to promote learners' employment of self-testing. Whether these interventions can be effective in high-stakes classroom or online learning is an important issue for future research. Importantly, research suggests that frequent low-stakes testing may be an effective method of reducing test anxiety. Although the testing effect is very general, testing can also have negative consequences, such as when choosing an incorrect answer in a multiple-choice test stamps that incorrect answer into memory and increases its likelihood of being recalled later. Understanding the conditions in which positive or negative consequences of testing are observed bears considerable importance regarding the theoretical understanding of test-enhanced learning. Characterizing, understanding, and exploiting the multifaceted effects of tests on long-term learning has provided a rich and deep challenge to researchers in psychology, education, cognitive science, neuroscience, and related fields.

Keywords: learning, metacognition, quizzes, retrieval, test anxiety, tests, test-enhanced learning, transfer

Subjects: Cognitive Psychology/Neuroscience, Educational/School Psychology

Introduction

According to a large-scale meta-analysis, the average effect size in rigorous randomized controlled trials of interventions designed to enhance educational outcomes is close to zero (Lortie-Forgues & Inglis, 2019). Against this disappointing backdrop, over a century of research has consistently demonstrated that undertaking practice tests (i.e., retrieving information from memory) is an effective strategy to boost learning and consolidate long-term memory. This article provides an up-to-date, selective review of key findings emerging from this body of research (for historical reviews, see McDermott, 2021; Roediger & Karpicke, 2006b).

The enhancing effect of testing is known by several names, including the *testing effect*, the *retrieval practice effect*, *retrieval-enhanced* and *test-enhanced learning*, and *test-potentiated learning*. It should be highlighted that this effect is not a single phenomenon but is composed of several distinct components driven, in all likelihood, by many underlying cognitive and motivational processes. For instance, it is well established that testing of studied information can consolidate its long-term retention, the *backward testing effect*. In addition, dozens of studies have found that testing of studied information can also facilitate subsequent learning of new information, the *forward testing effect* or *test-potentiated new learning*. Furthermore, studies have even observed that engaging in a test on to-be-learned information (i.e., prior to exposure to the learning content) can boost subsequent learning, the *pretesting* or *pre-question effect*. Characterizing, understanding, and exploiting the multifaceted effects of tests on long-term learning have provided a rich and deep challenge to researchers in psychology, education, cognitive science, neuroscience, and related fields.

This article summarizes key findings on the major components of test-enhanced learning and discusses practical (under)employment of testing in educational settings. Some of the evidence for both the generality of test-enhanced learning and its limitations is reviewed, including contexts in which tests have detrimental effects. Methods for optimizing the benefits of tests are discussed, and some of the major theoretical ideas that have been proposed to explain why tests have such powerful effects in facilitating and consolidating learning are considered. Table 1 provides a glossary of key terms.

Empirical Findings on Test-Enhanced Learning

Backward Testing Effect

The act of retrieving information from memory serves to strengthen that memory and/or make it more accessible. A classic example of this backward testing effect was provided by Roediger and Karpicke (2006a), who instructed college students to study text passages on scientific topics. During the learning phase, participants either repeatedly studied a passage four times (SSSS), studied it three times and then took a practice recall test (SSST) or studied it once and completed three practice tests (STTT). In the practice tests, participants freely recalled as much information as they could from the studied passage, without any feedback about their accuracy. This test format, termed free recall, is similar to an essay test in a real classroom. One week later, participants took a final free recall test on the studied text. The results showed that final test performance systematically increased as a function of the number of practice tests, with performance at 40% accuracy in the SSSS group, 56% in the SSST group, and 61% in the STTT group, implying that, by comparison with restudying, testing more effectively consolidates long-term retention of studied information. Strikingly, the benefit of repeated practice tests in the STTT group was sufficiently strong to overcome the fact that four times as much exposure to the text was given to participants in the SSSS group. (When the test was administered 5 minutes

rather than 1 week after the learning stage, the SSSS group outperformed the groups that had practiced retrieval. This issue is further discussed in the section on "Negative Consequences of Tests.")

Table 1. Glossary of Key Terms and Abbreviations

Backward testing effect	The classic testing effect (e.g., Gates, 1917; Spitzer, 1939) in which practicing retrieval of studied information increases later recall, relative to a suitable control treatment (typically with restudying replacing retrieval).
Forward testing effect	A form of testing effect (e.g., McDermott, 2021; Szpunar et al., 2008; Tulving & Watkins, 1974) in which practicing retrieval of studied information facilitates learning of subsequent information. Also known as *test-enhanced* or *test-potentiated new learning*.
Grain size effect	Smaller chunks of learning (e.g., shorter text passages, shorter word lists), each followed by a practice test, should increase retrieval success during those tests and in turn enhance subsequent memory, compared to larger chunks.
Pretesting effect	Engaging in a test on to-be-learned information (prior to exposure to the learning materials) can boost subsequent learning. Also known as the *pre-question effect*.
Retrieval practice	Taking a test and hence practicing retrieval (e.g., cued or free recall) of studied information.
Test-enhanced learning	General term for all benefits caused by testing, including the backward and forward testing effect, but also indirect effects such as enhanced learning resulting from alleviation of test anxiety, reduction of mind-wandering, and identification of knowledge gaps.

Testing also tends to be more powerful than other elaborative study strategies, such as note-taking (Heitmann et al., 2018) and concept-mapping (Karpicke & Blunt, 2011). For instance, Karpicke and Blunt (2011) asked four groups of participants to study a science text in preparation for a short answer test administered a week later. A concept-mapping group studied the text once and then constructed a concept map of the concepts described in the text. A retrieval-practice group studied the text once, took a practice free recall test, and then repeated this study-recall cycle once more. Lastly, a study-once group studied the text only once, while a repeated-study group successively studied the text four times. It is worth noting is that the total amount of learning time was exactly matched between the concept-mapping and retrieval-practice groups. After the learning session, all four groups predicted the percentage of information they would remember in the test. These subjective predictions were collected to assess participants' metacognitive awareness about the effectiveness of different study strategies (rereading vs. concept-mapping vs. retrieval practice). One week later, all four groups returned to the laboratory to take a final test, which included both verbatim and inference questions. The results showed that final test performance was better in the retrieval-practice group than in the other three

groups, regardless of whether the question type was verbatim or inferential. Intriguingly, in contrast to objective test performance, participants' subjective predictions demonstrated the reverse pattern: predictions were lower in the retrieval-practice than in the other three groups, implying that learners can lack metacognitive awareness of test-enhanced learning. The topic of metacognition is discussed in more detail in the section "(Under)Employment of Testing to Enhance Learning."

Test-enhanced consolidation of studied information has also been documented in several meta-analyses (e.g., Adesope et al., 2017; Rowland, 2014; Yang et al., 2021). For instance, Rowland (2014) integrated results across 61 laboratory studies to compare the effectiveness of testing and restudying. The results again showed that testing was more effective than restudying (Hedges' g = 0.50), an effect that is equivalent to moving from the 50th to the 69th percentile of a normal distribution. Importantly, these meta-analyses found that the backward testing effect generalized to different types of study materials (including prose passages, paired associates, and single words), to different test formats (including cued recall, free recall, and recognition), across a wide range of time intervals between study and practice test and between study and final test, regardless of whether corrective feedback (i.e., correct answers) was provided in the practice tests, and occurred both when test format was matched and mismatched between practice tests and final assessments (see the section "How General Are the Benefits of Retrieval-Enhanced Learning?" for more discussion about these factors).

Finally, it is important to note that although there are some indications of publication bias—the selective publication of statistically significant results and the censoring of non-significant effects—in this field (as in virtually all fields in the behavioral sciences), the major phenomena related to test-enhanced learning remain highly robust after correction for publication bias (Boustani & Shanks, 2022; Yang et al., 2021), and demonstrations of testing effects in high-powered, pre-registered experiments are beginning to be published (e.g., Don et al., 2022).

In summary, testing is a more powerful strategy to consolidate retention of studied information than restudying and other elaborative strategies (e.g., concept-mapping). In addition, the backward testing effect is generalizable to different test formats, study-test intervals, and material types.

Forward Testing Effect

Besides consolidating retention of studied information, practice tests can also prospectively facilitate subsequent learning of new information. Szpunar et al. (2013) provided a clear demonstration of the forward effect of testing. In their study, two groups of undergraduates watched a four-segment lecture video on statistics concepts, having been instructed that there would be a final recall test on the contents of all segments. After watching each of Segments 1–3, a restudy group restudied the six key concepts delivered in the just-studied segment, whereas a test group undertook a short answer test on these concepts. After completion of Segment 3, both

groups watched Segment 4 and engaged in a short answer test on this segment. The results showed a clear forward testing effect: Testing on Segments 1–3, by comparison with restudying them, substantially improved test performance on Segment 4 (test: 89% vs. restudy: 65%).

Similar to the backward testing effect, the forward testing effect is also generalizable to different types of study materials (including word lists, paired associates, text passages, lecture videos, artists' painting styles) and populations (including elementary school children, college students with different levels of working memory capacity and test anxiety, older adults, and patients with traumatic brain injury; for reviews, see Pastötter & Bäuml, 2014; Yang et al., 2018).

Pretesting Effect

In contrast to the forward and backward testing effects, the pretesting effect refers to the fact that undertaking a test before exposure to learning content can facilitate subsequent learning of that content. For instance, Carpenter and Toftness (2017) instructed two groups of participants to watch an educational video composed of three segments about the history of Easter Island. Prior to viewing each section, a pre-question group answered two questions about the content that would be delivered in the next segment while a control group did not answer these pre-questions. After viewing all three segments, both groups took a final test, which included 12 test questions. Among the 12 questions, six were the same as used in the pretests, and the other six were new questions (not used in the pretests). These new questions were introduced to explore whether pretesting can also enhance learning of untested (non-pre-questioned) information. Carpenter and Toftness's results showed that pretesting not only facilitated subsequent learning of tested information but also of untested information.

Although these findings are encouraging in showing that pretesting facilitates learning of both tested and untested information, it should be noted that other studies have observed inconsistent findings by showing that pretesting promotes learning of tested information but does not benefit (Carpenter et al., 2018; Latimier et al., 2019) or even impairs (Peeck, 1970; Sagaria & Di Vesta, 1978) learning of untested information. Another noteworthy point is that Latimier et al. (2019) recently reported that the enhancing magnitude of post-testing (i.e., administering tests after studying; Cohen's d = 0.74) was more than double that of pretesting (d = 0.35). A similar finding was observed in the meta-analysis of Yang et al. (2021), which found that the enhancing effect of post-testing (Hedges' g = 0.54) is much greater than that of pretesting (g = 0.19), although both effects were statistically reliable. These findings suggest that post-testing is relatively more effective than pretesting.

Overall, pretesting can promote subsequent learning of tested information, but its effect on non-pre-questioned information remains unclear. The enhancing effect of pretesting is relatively smaller than the effect of post-testing, implying that practitioners should primarily consider utilizing post-testing.

Test-Enhanced Learning in the Classroom

The empirical studies described previously were all conducted in the laboratory, and readers may wonder whether test-enhanced learning generalizes to real classroom settings. The answer to this question is affirmative. Hundreds of quasi-experiments have documented that frequent quizzing accelerates classroom learning (for reviews, see Schwieren et al., 2017; Trumbo et al., 2021; Yang et al., 2021). For instance, Roediger et al. (2011) explored the effect of low-stakes multiple-choice quizzes on classroom learning. In this study, middle school students attended a social studies course, with half of the course content regularly quizzed in a multiple-choice format, and the other half unquizzed. Students' course knowledge was assessed by chapter and semester exams. The results showed that, in both chapter (quizzed: 94% vs. unquizzed: 81%) and semester (quizzed: 79% vs. unquizzed: 67%) exams, students' exam performance was better for quizzed than for unquizzed course content.

Yang et al. (2021) conducted a meta-analysis to explore if quizzing boosts classroom learning. The results showed a medium-sized enhancing effect of quizzing on students' attainment (g = 0.50). In summary, the enhancing effect of testing survives in the classroom, and frequent quizzing can be employed as an efficient tool to facilitate teaching and learning in real educational settings.

(Under)Employment of Testing to Enhance Learning

Even though numerous studies have established the power of testing for enhancing learning and consolidating memory, it is striking that, in real educational situations, testing seems not to be as widely used as it deserves to be. A possible explanation for underemployment of retrieval practice is that learners lack metacognitive appreciation of the benefits of testing (for a review, see Rivers, 2021). As discussed in the "Backward Testing Effect" section, Karpicke and Blunt (2011) observed that, although participants actually memorized tested texts better, they provided lower test performance predictions for tested than for reread texts and for ones studied with concept-mapping (for related findings, see Roediger & Karpicke, 2006a). Consistent with this, research has tended to paint a rather unflattering picture of the average learner's metacognitive understanding of the value of retrieval practice. However, a slightly more nuanced and positive perspective has emerged as research on this topic has accumulated.

In a survey study conducted by Karpicke et al. (2009), university students were asked to report study strategies that they frequently used and to rank the reported strategies in the order of usage frequency. Among these students, a majority (83.6%) reported rereading (i.e., restudying), and only a minority (10.7%) mentioned practice retrieval. Furthermore, 55% of them identified rereading as their most frequently used strategy, and only 1% indicated that practice retrieval was the one they employed most frequently.

Similar findings were observed in questionnaire studies presenting vignettes of learning scenarios (Karpicke et al., 2009; McAndrew et al., 2016; McCabe, 2011; Morehead et al., 2016). In these vignettes, participants were informed that hypothetical students were preparing for an exam, and that they either tested themselves or restudied the course materials. Participants were

required to indicate which of the two strategies was more effective. These survey studies have consistently found that most participants regard restudying as more effective, with only a minority endorsing testing as more effective. Poor metacognitive calibration is not confined to just the backward testing effect. Lee and Ahn (2018) measured metacognitive judgments of learning in the forward testing effect. Participants in the interim test condition showed better performance, and higher metacognitive judgments, but they still underestimated their performance.

In a review of metacognition in the testing effect, Rivers (2021) combined results across 10 questionnaire studies (including results from 4,240 participants). These questionnaire studies asked participants to report the primary reason why they tested themselves when studying. The aggregated results showed that a majority (52%) of participants tested themselves to determine how well they have learned the study materials, and only 26% reported that they would learn better through testing than through rereading. These findings suggest that learners typically regard testing as an assessment tool to measure learning progress, rather than an efficient strategy to accelerate learning.

Other research, however, suggests that learners' understanding and employment of testing might be somewhat richer than implied by the studies described. Kuhbandner and Emmerdinger (2019) noted that while rereading is indeed usually an ineffective study strategy, surveys typically have not asked about "rereading" but about "restudying," a term that respondents might interpret in a range of ways, including "rereading not understood text," a perfectly sensible study method. They also noted that surveys have rarely given respondents the opportunity to distinguish their study strategies early versus late in learning. Kuhbandner and Emmerdinger provided empirical evidence that higher rates of self-testing are reported when ambiguity about the term "restudying" is eliminated and respondents are able to describe their strategies both early and late during a university course. It is also quite likely that learners understand the value of retrieval practice, even if they do not employ it to a level that would maximize their exam attainment (Blasiman et al., 2017). Lack of motivation or insufficient knowledge about how best to exploit testing may be factors that suppress its use.

Improving Control via Experience

These findings show two things: first, when study method is imposed by the experimenter, participants often do not appreciate how effective retrieval practice is for their own long-term learning; and second, when surveyed, only a minority of students report employing retrieval practice to consolidate their learning. But what happens when students actually have control over their study method in an experimental learning task? Are they more likely to employ retrieval practice under such conditions? Even here, misjudgment of the benefits of retrieval practice impacts how often learners use retrieval practice as a learning strategy. Karpicke (2009) had participants learn foreign language word-pairs, and word pairs could be tested, restudied, or dropped from study. When learning strategies were experimenter-controlled, repeated retrieval practice following one successful recall improved retention more than repeated restudy or dropping the item from further practice. However, when learning strategies were self-regulated,

participants did not practice retrieval as early or as often as optimal, which subsequently impaired their performance. Instead, participants tended to continue to restudy pairs, particularly early in training, and dropped pairs from study after only one successful recall. Thus, learners do not always use retrieval practice effectively when self-regulating learning.

More encouragingly, self-testing is generally preferred when the likelihood of successful retrieval is high. That is, when the tests are easy, when the material is easy, well-learned, or has been studied recently, or when feedback or an opportunity to restudy is provided during testing, employment of self-testing increases (e.g., Persky, 2018; Toppino et al., 2018; Tullis et al., 2018; Vaughn & Kornell, 2019).

One question is whether people can appreciate through experience that retrieval practice is better for learning than restudying. Some studies have shown that learners are able to adaptively modify their strategies through experience with the nature of study materials and tests (e.g., Finley et al., 2010) and therefore it is possible that simple experience with the benefits of testing may allow learners to adapt their beliefs and strategies. However, Tullis et al. (2013) found that participants altered their beliefs only when they were provided with additional external support, for instance, summaries of performance for tested and studied material. In addition, this change in beliefs occurred for only half the participants. In contrast, Higham et al. (2022) found that judgments of learning improved over repeated sessions, indirectly suggesting that use of testing might change as a consequence. Thus, the evidence is mixed on whether metacognitive knowledge about retrieval practice can improve through experience alone. Hui et al. (2021) investigated learners' strategy use after receiving individual feedback about their own learning outcomes. Participants received feedback on their recall of items they had restudied or self-tested. Participants who experienced the testing effect rated retrieval practice as more effective, and also chose retrieval practice as a learning strategy in a new learning phase, to a greater extent than those who did not experience a benefit.

Improving Control via Strategy Instructions

Although the findings are rather disappointing, in that learners lack metacognitive awareness of the benefits of testing and do not frequently use testing to enhance their learning, more encouraging is that researchers have developed several effective interventions to promote practical use of self-testing, such as direct instructions about the beneficial effects of testing (Ariel & Karpicke, 2018; McCabe, 2011) and creating contexts that support self-testing (Bottiroli et al., 2010; McCabe & Lummis, 2018; Vaughn & Kornell, 2019). For instance, Ariel and Karpicke (2018) investigated self-regulated learning of foreign language translations, where participants could study, self-test, or drop items from learning across multiple practice blocks. Participants were given either neutral instructions or detailed instructions about the benefits of retrieval practice, and how to use it to maximize retention, specifically stating that retrieval should be repeated three times before items are dropped from study. In the strategy instruction group, participants were less likely to terminate learning for items after an initial successful retrieval and continued to use a repeated retrieval strategy in a second session a week later when learning new material. Items that were retrieved successfully multiple times during practice were more

likely be recalled in the final test. Broeren et al. (2021) used a similar instructional intervention in an online higher education course with complex materials. They found that the strategy instruction group used testing as a study strategy more often and had a greater number of correct retrieval attempts during the study sessions. However, in this case, there was no effect of instructions in the final course exam, or correlation between use of retrieval practice and exam performance. Hui et al. (2021), in the research described previously, additionally found that general strategy instructions were ineffective for participants who experienced better recall for restudied than self-tested items, or the same recall level for both. These participants instead preferred to continue restudying.

Overall, although the beneficial effects of testing on learning and memory are substantial, learners tend to underappreciate the merits of practice tests, leading to their underemployment. However, some promising interventions have been developed to promote learners' employment of self-testing. Whether these interventions can be effective in high-stakes classroom or online learning is an important issue for future research.

Potential Concerns About Applying Retrieval-Enhanced Learning

The sheer number of studies demonstrating retrieval-enhanced learning raises the important question about both its limits and the factors that moderate its magnitude. This section starts by reviewing evidence that the effect has extraordinary generality before describing conditions where testing is not advantageous or indeed even has negative effects.

How General Are the Benefits of Retrieval-Enhanced Learning?

The meta-analysis by Yang et al. (2021) integrated data from nearly 50,000 students extracted from 222 independent studies. This rich resource allowed a number of variables to be examined. To begin with, several variables are described that appear to have very little moderating role on the benefits of testing and which therefore confirm its generality. First, Yang et al. found that quizzes administered out of the classroom significantly boost students' attainment, although not quite as effectively as ones administered in class. Hence, they generalize across test context. This is encouraging for teachers who may be concerned about the loss of in-class teaching time resulting from administering frequent tests.

New technologies, collaborative learning platforms (e.g., PeerWise; see Kay et al., 2020), and smartphones are available to make it easy to administer and score quizzes. These techniques are virtually cost-free and allow teachers to easily implement immediate quizzes. The meta-analysis by Yang et al. (2021) found that the quiz administration mode (e.g., paper-and-pen, web-based, oral, or clicker response system) has minimal influence on the magnitude of test-enhanced learning.

The meta-analysis found that all kinds of test formats (e.g., matching, fill-in-the-blank, short answer, multiple-choice, cued recall, free recall, and so on) reliably enhance learning; the testing effect generalizes across all subject categories for which sufficient data are available (e.g.,

biology, chemistry, geography, history, language/reading/vocabulary, psychology, and so on). Test-enhanced learning is effective across different education levels, including elementary school, middle school, high school, and university/college. These findings document the stability and generalizability of test-enhanced learning to different test formats, subjects, and populations.

In their meta-analysis, Yang et al. (2021) coded classroom studies into three categories according to their exam content type: Fact (i.e., memory of specific course contents, such as historical events), Concept (e.g., concept comprehension and inductive inferences that go beyond direct experience, such as learning rules of syntax in a foreign language course), and Problem-solving (i.e., applying learned knowledge or skills to solve problems, such as performing resuscitation treatment in cardiac arrest scenarios). Yang et al. found that tests not only promote learning of facts, but also facilitate concept learning (Jacoby et al., 2010; Karpicke et al., 2014; Yang & Shanks, 2018) and enhance knowledge application in the service of problem-solving (Kromann et al., 2010; Sennhenn-Kirchner et al., 2018). Thus, retrieval practice can enhance learning at multiple levels of Bloom's taxonomy (Krathwohl, 2002).

While these results are encouraging regarding the generality of test-enhanced learning, they come from a meta-analysis in which the results of many diverse studies were integrated, and in addition they are negative results—there was little indication that the testing effect differs in magnitude across the factors discussed here (such as the content type and test format). They should therefore be interpreted as suggestive rather than conclusive.

Does frequent quizzing widen the achievement gap (e.g., between novices and experts in a subject area, less-skilled and more-skilled readers, students with overall lower and higher academic performance)? Do some students benefit less from quizzes or do quizzes exert detrimental effects on some students' academic performance because they typically perform less well on assessments? Very little research has addressed this important question, but Yang et al. (2020) observed that the benefits of testing are, if anything, larger for individuals with low than with high working memory capacity.

Does Frequent Quizzing Provoke or Alleviate Test Anxiety?

A very legitimate concern is that quizzing may provoke test anxiety. Several lines of evidence are relevant to this important issue.

Low-stake (or no-stake) tests arouse minimal test anxiety, so it is straightforward to ask whether they can be shown to effectively promote learning. Yang et al. (2021) divided 295 classroom experiments into two subcategories (high/low) according to the stake-level (i.e., whether quiz performance was incorporated into course grades). Both high- and low-stake quizzes significantly enhanced students' exam performance, and there was no significant difference between them, revealing that even low- or no-stake quizzes can facilitate learning.

Importantly, frequent testing has little impact on or even reduces (rather than increases) test anxiety. For instance, in a large sample study (over 1,000 college participants), Yang et al. (2020) observed that interpolating tests across a study phase has minimal influence on participants' test anxiety. Szpunar et al. (2013) found that frequent tests significantly relieve test anxiety (for related findings, see Higham et al., 2022; Khanna, 2015). Furthermore, in a large-scale survey conducted by Agarwal et al. (2014), 72% of 1,306 middle and 102 high school students reported that frequent quizzes made them less anxious about exams, with only 8% reporting the opposite. In another survey conducted by Sullivan (2017), about 91% of 353 college students agreed with the statement "*The option to retake a quiz reduced my test anxiety*," and only 3.1% disagreed.

Although the number of studies on the impact of the testing effect on test anxiety is modest, Yang et al. (2023) were able to identify 25 effects based on data from over 3,000 participants. When submitted to a meta-analysis, these data revealed a clear negative effect of frequent tests on test anxiety (Hedges' g = -0.626): tests reliably reduce anxiety. This beneficial effect was not moderated by setting (laboratory vs. classroom), the format in which tests were administered (e.g., recall, short answer, multiple choice), the nature of the control condition, the population, or the academic subject. Again, these null results regarding moderators should be interpreted with some caution, but there seems fairly clear evidence that the alleviating effect of tests on test anxiety is wide and quite robust.

Negative Consequences of Tests

While the findings reviewed here seem to suggest that the benefits of testing on learning are extremely general, it would be a mistake to infer that negative consequences are never observed. On the contrary, such negative effects are well-established and have considerable importance regarding the theoretical understanding of test-enhanced learning.

One relatively straightforward set of conditions in which detrimental effects of testing are observed, relative to restudying, is when the retention interval between study and the final test is short, on the order of a few minutes (Roediger & Karpicke, 2006a). Even here, though, several features of the situation (including absence of feedback in the practice test and low test performance) seem to be necessary to evoke this pattern (Rowland & DeLosh, 2015). When no feedback is provided, then any unrecalled items in the practice test will fail to gain any learning boost, whereas all items receive a re-exposure boost in the restudy condition. Even though the latter may be short-lived, it can be sufficient to enhance retention above the level of the retrieval practice group when the retention interval is short.

Perhaps more interesting are negative effects that arise when study and test trials are interleaved. Finn and Roediger (2013) devised an innovative experimental design in which participants initially studied a set of face-name pairs. Next, they learned new information comprising the professions associated with these faces but did so immediately after either restudying the relevant face-name pair (the restudy condition) or retrieving the name given the face (the retrieval condition—corrective feedback was given on trials where recall failed). Finally, recall of all names and professions was attempted given the faces. Finn and Roediger found that while

name retrieval enhanced final recall of the face-name associations relative to the restudy control condition, it impaired recall of the face-profession associations. In a later replication of the effect, Davis et al. (2017) presented some evidence that frequent task-switching between retrieval (of the face-name association) and new learning (the face-profession association) is part of the cause of the effect. In conditions where all face-name associations are retrieved prior to new learning of all the face-profession associations, the detrimental effect of retrieval is not found.

The well-known phenomenon of retrieval-enhanced forgetting provides another example of the harmful consequences on one set of memory items when another set is retrieved. In a common method for obtaining this well-established phenomenon (see Murayama et al., 2014, for a meta-analysis), participants study a set of category-exemplar pairs such as *fruit-lemon*, *drinks-gin*, *fruit-banana*, *drinks-whiskey*. In the retrieval practice stage, they attempt to recall some of the exemplars from some of the categories (e.g., *fruit-le*____). As expected, this retrieval practice boosts recall of these items (e.g., *lemon*) in a final recall test of all the exemplars, a clear example of test-enhanced learning. At the same time, however, recall of the untested items (e.g., *banana*) becomes suppressed, not only in comparison with the practiced items but also relative to appropriate control exemplars from nonstudied categories. This retrieval-enhanced forgetting has been documented with a very wide range of materials and experimental procedures.

There is much debate about the mechanism or mechanisms that create this effect (see Murayama et al., 2014). From the perspective of delineating the boundaries of test-enhanced learning, the evidence suggests that retrieval-enhanced forgetting is best viewed as a temporary reduction in the accessibility of non-retrieved items in memory (Storm & Soares, 2022). Indeed, not only is the effect often short-lived, but there is abundant other evidence that retrieval of particular elements of learned information can boost rather than impair retention of untested elements. The meta-analysis of Yang et al. (2021) found that testing significantly benefits untested knowledge, although to a smaller extent (g = 0.321) than for tested knowledge (g = 0.512), and a similar result was obtained in an earlier meta-analysis by Pan and Rickard (2018). Like retrieval-enhanced forgetting, this facilitation effect occurs with a wide range of materials (see Bäuml & Trißl, 2022). As a concrete example, Yang et al. (2019) found that retrieval practice on factual information about artists (e.g., *Veronese introduced a greater realism and sumptuous, decorative color*) induced a forward testing effect on subsequent visual learning of artists' styles. Chan (2009) discussed some of the factors that determine whether retrieval induces forgetting or facilitation.

Negative testing effects have been demonstrated in other contexts, including susceptibility to misinformation (Chan et al., 2009, 2022), and in experiments where retrieval disrupts organizational processing (Mulligan, 2020).

The final topic in this section involves tests that employ multiple-choice questions (MCQs). The possibility of negative testing effects with such tests has been extensively studied, given the intuition that choosing an incorrect MCQ answer might stamp that incorrect answer into memory and increase its likelihood of being recalled later. Indeed, such a pattern was demonstrated by

Roediger and Marsh (2005). This effect is heavily dependent on corrective feedback being withheld in the practice MCQ test: when feedback is provided, the negative effect of selecting a lure in the practice test is largely eliminated (Butler & Roediger, 2008).

Several studies have examined situations where a lure in one MCQ is the correct answer in a later one (Little et al., 2012, were the first to examine this issue). For instance, suppose that in an initial MCQ test, an item is "What is the name of the large bone in the upper arm?" and the answer options include *humerus* (correct) and *femur* and *tibia* (lures). Then in a final test, a related MCQ asks "What is the name of the large bone in the upper leg?" and the answer options include *femur* (correct) and *humerus* and *tibia* (lures). Alamri and Higham (2022) showed that when a lure (e.g., *tibia*) was incorrectly chosen in the initial test, the corrective feedback (in this case *humerus*) received a memory boost that caused it to be incorrectly selected much more often in the related MCQ than would be the case if the questions were unrelated. Two aspects of these results are noteworthy. First, the effect concerns choice of the feedback item (*humerus*), not the lure the participant incorrectly chose (*tibia*); participants seem to be swayed by the enhanced familiarity of the answer they presumably processed deeply as corrective feedback, failing to appreciate that the new question is different from the initial one. Secondly, these subtle effects must be viewed in the context of an overall benefit of MCQ testing, both when the practice MCQ involves the same items as the final test (the pattern observed by Roediger & Marsh, 2005) and when the items are different but related (see Little, 2018; Little et al., 2012). The finding that taking an MCQ test can enhance performance on a later test with related materials provides yet another example of the benefits of testing generalizing beyond the tested materials.

Optimizing the Benefits of Retrieval

How can retrieval practice be leveraged to provide the best benefits for long-term retrieval? Research has suggested a number of answers to this important question and highlighted several contributing factors.

Role of Feedback

Retrieval practice promotes long term retention, even in the absence of feedback. However, successful retrieval at practice is critical in learning from tests: successful retrieval attempts implicitly provide self-generated feedback to the learner, and (unless practice retrieval is perfect) providing the correct answer in the form of feedback increases the benefits of retrieval practice (Bangert-Drowns et al., 1991; Kulhavy & Stock, 1989), as it allows correction of errors (Pashler et al., 2005) and maintenance of correct responses (Butler et al., 2008). Thus, increasing the success of retrieval practice through feedback benefits long-term retention, and has few, if any, disadvantages (Erdman & Chan, 2013).

Some studies found that testing effects were reversed (restudy performed better than test) with repetitive feedback in delayed tests following an initial test with feedback, unless feedback was also received during training (Pastötter & Bäuml, 2016; Storm et al., 2014). However, retrieval success in the test conditions during practice was poor and was improved by providing feedback

during practice. Racsmány et al. (2020) found that simply presenting items multiple times prior to practice is sufficient to protect against reversal of the testing effect with feedback. Thus, protection from reversal of the testing effect was likely due to an increase in retrieval success at practice.

Research on the optimal timing of feedback has provided conflicting results. The popular view is that feedback should be provided immediately (e.g., Hattie & Timperley, 2007), deriving from behaviorist theories that immediate feedback is critical to correct errors and reinforce correct responses. In a meta-analysis, Kulik and Kulik (1988) suggested that immediate feedback is more effective in applied classroom studies, while delayed feedback is more effective in laboratory studies. However, there is growing evidence for a general benefit of delayed over immediate feedback. Butler et al. (2007) assessed the effect of the timing of feedback on learning and recall. Participants studied prose passages, each followed by a multiple-choice test. Feedback was presented immediately or after a delay of 10 minutes (Experiment 1), or after 24 hours (Experiment 2). In a final test, performance was better for items that had received delayed feedback than for ones that had received immediate feedback. In addition, Mullet et al. (2014) found that delayed feedback on homework assignments improved retention and transfer of learning in a university engineering course, although students preferred immediate feedback (see also Lefevre & Cox, 2017).

It has been argued that delayed feedback may benefit learning, as it reduces interference from errors when learning correct responses (Kulhavy, 1977), or through the benefits of spaced over massed presentation (see Cepeda et al., 2006). The effects of feedback on test-potentiated new learning are less clear. A meta-analysis has shown that the forward testing effect occurs reliably in the absence of feedback and is reduced rather than enhanced when feedback is provided (Chan, Meissner, & Davis, 2018). While this result contrasts with the substantial benefits of feedback in the backward testing effect, this analysis was based on a limited number of effect sizes, and was confounded with effects of interleaved testing procedures (Boustani & Shanks, 2022). It therefore warrants further investigation.

Optimal Placement of Tests

What is the optimal placement of retrieval practice to promote retention? The grain size hypothesis states that interspersing tests of smaller chunks of information throughout learning will benefit long-term retention compared to testing larger chunks of information at the end of learning. Interim tests might facilitate better retention due to improved retrieval success during practice tests (Lavigne & Risko, 2018; Weinstein et al., 2016), which is fundamental to long-term retention (Pyc & Rawson, 2009) and might be a necessary condition for sufficient elaborative retrieval (Rawson et al., 2015). In addition, interim testing during a lecture can increase students' employment of positive study behaviors such as note taking (Szpunar et al., 2013), reduce overconfidence (Szpunar et al., 2014), and improve integration of content across sections (Jing et al., 2016). Healy et al. (2017) have argued that interim tests might promote learning in line with

the *cognitive antidote principle*, which suggests that making a monotonous task more difficult increases attention and performance. As described, in many contexts interim tests potentiate new learning of subsequent information, which might foster better overall retention.

However, while interim tests often enhance performance during practice, most research has quite surprisingly found little evidence that this translates to a benefit in recall in a final test. Wissman and Rawson (2015) conducted seven experiments assessing whether there was a difference between interim and end test schedules in remembering lengthy expository texts, broken into sections. Participants completed either free recall following the study of each section (interim test condition) or free recall following study of all sections (end test condition). After a 20-minute retention interval, participants completed a final 20-minute free recall test. Despite consistent benefits of interim tests compared to end tests across practice, this was not maintained in the final test. The absence of a grain-size effect in final recall has been further replicated in the laboratory (Latimier et al., 2020; Uner & Roediger, 2018) and the classroom (Weinstein et al., 2016). Latimier et al. (2020) compared small, medium, and large grain size schedules for re-reading and testing and found that while placement was important for re-reading, with greater benefits of interspersed re-reading than postponed re-reading, test placement made no difference for retrieval practice. In one study that did find an effect of grain size on final recall (Healy et al., 2017), there was an interaction between list order and grain size, with a benefit of interim over end tests only for later lists. Here, participants completing interim tests reported significantly more engagement with material in later lists, which likely contributed toward improved retention for later lists.

While grain-size effects appear elusive in complex texts, research (Boustani et al., 2022) has demonstrated significant grain-size effects with word lists, although there was still a substantial loss of retention between practice and final test in these experiments. This difference in findings for lengthy texts and simple word lists merits further research.

Although involving the optimal form rather than placement of tests, a related question regarding test-potentiated new learning is whether it is necessary for the interim tests administered after each study block to test all information presented in that block. The clear answer from one set of experiments is "no." Using face-name and foreign language translations, Don et al. (2022) found that tests potentiated new learning even when they tested only a fraction of the items studied in the preceding block. Thus, there is some reason to believe that the effects of retrieval practice are optimized when they are partial, freeing up time for other beneficial learning activities.

Theoretical Accounts of Test-Enhanced Learning

As noted in the Introduction, test-enhanced learning is not a unitary concept reducible to a single mechanism but rather a diverse set of phenomena caused, in all likelihood, by a diverse set of underlying processes. With this context in mind, it should be little surprise that a broad range of theoretical accounts has been developed. Evaluation of these theories should take due heed of the possibility that they may not be mutually exclusive, and each account may be valid in a specific set of circumstances.

Elaborative Retrieval Hypothesis

Historically, the most dominant theories of the testing effect have been based on the plausible idea that retrieval facilitates semantic elaboration (for a review, see Karpicke, 2017). One influential version of this account, the *elaborative retrieval hypothesis* (Carpenter, 2009) conjectures that when a target piece of information is retrieved from memory, such as a word or text, there is spreading activation of information semantically related to the target. This benefits memory, as the related information can be used as a semantic cue increasing the number of retrieval pathways leading to the target. A more specific variant of this model has been proposed for words and word-pairs, the *mediator effectiveness hypothesis* (Pyc & Rawson, 2010). Similarly, this account states that the testing effect in paired-associate learning is due to the activation of semantically related words, which can then be used as mediators to recall the target. For example, take the cue-target word-pair *mother-child*: in the testing condition participants are provided only the cue, *mother*, and are asked to recall the target, *child* (in contrast, in the restudy condition participants are provided the word-pair again to study). According to the mediator hypothesis, attempting to retrieve the target word leads to semantic elaboration, and hence activation of semantically related words, such as *father*, which is a strong forward associate of the cue, *mother*, and which can then function as a mediator. If the target is successfully recalled, then the mediator-target association becomes strengthened such that the mediator can later function as an additional cue to facilitate target retrieval. That is, both the cue, *mother*, and mediator, *father*, can be used to recall the target, *child*. As restudy does not involve either retrieval or semantic elaboration, testing produces better learning.

The elaborative retrieval and mediator effectiveness hypotheses are founded on two key assumptions: that semantic elaboration occurs more in retrieval than in restudy, and that semantic elaboration benefits learning. On both counts, the evidence is inconsistent. Although some research on spontaneous mediator generation has found that mediator cues are more strongly activated by testing than restudy (Carpenter, 2011; Pyc & Rawson, 2010), this effect is not ubiquitous, and other research has found no significant differences between restudy and testing in mediator activation (Cho et al., 2017; Lehman & Karpicke, 2016). Similarly, research addressing whether mediator cue generation and semantic elaboration enhance learning has also been inconsistent. Some studies have found that the strength of mediating cues is important for the testing effect (e.g., Carpenter, 2009; Pyc & Rawson, 2010), but others have found no interaction (Cho et al., 2017; Lehman & Karpicke, 2016). Likewise, the benefit of elaboration on learning is also disputed. Although some research has demonstrated that elaborative prompts provided during testing can enhance learning (Endres et al., 2017), there are many examples where this is not the case and tasks that should increase elaboration have not benefited learning (e.g., Whiffen & Karpicke, 2017).

These inconsistencies do not discredit elaborative and mediator hypotheses, which have proved useful for explaining the testing effects found with semantically related materials. However, they do suggest that there are significant limitations to the generalizability of the explanations to all contexts. There are many cases where a testing effect is produced, but semantic elaboration is either unlikely or irrelevant, for example, in learning symbols (Coppens et al., 2011), visuospatial information (Kang, 2010), and maps (Carpenter & Pashler, 2007).

Episodic Context Account

The limitations of the elaborative retrieval hypothesis led to the development of the episodic context account (Lehman et al., 2014). This account argues that initial retrieval causes the reinstatement of the learning context, and this memory is updated to include episodic (spatial-temporal) contextual cues from the current learning environment. This makes the memory more distinctive and accessible during subsequent recall attempts. For example, according to the episodic context account, attempting to retrieve the target word *child* when provided the cue *mother* does not lead to semantic elaboration but to remembering the details of the original learning episode, namely the spatial-temporal details of the learning occasion, which were bound to the target word. Reinstating the learning context leads to additional episodic cues from the current learning occasion binding to the target, making it more mnemonically distinct.

The episodic context account has received empirical support. Whiffen and Karpicke (2017) found that list discrimination tasks, which ask participants to recall list order, but not pleasantness ratings or category judgment tasks, produced a testing effect. This effect was also correlated with organization of recall output, where list discrimination tasks promoted temporal organization of recall output, whereas category judgment tasks promoted semantic organization. Preliminary research has also indicated that the testing effect is sensitive to contextual cues (Schwoebel et al., 2018) and that successful retrieval attempts are marked by neural areas correlated with processing of temporal-spatial details both at practice and test (Bai et al., 2015).

Other Theories

Retrieval Effort

One set of explanations for the testing effect argues that the increased difficulty of retrieval practice, relative to restudy, is directly related to its benefit on memory (Rowland, 2014). These theories can be conceptualized under the "desirable difficulties" framework of learning, which describes how difficult tasks requiring greater effort to complete result in better long-term retention (Bjork, 1994). Within this framework, retrieval practice represents a desirable difficulty, which, compared to the relative simplicity of rereading, strengthens memory of the retrieved content. In the new theory of disuse from Bjork and Bjork (1992), this may be due to the successful retrieval of an item increasing its storage and retrieval strength—that is, how well learned an item is and its momentary accessibility from memory. An important point to note is that although retrieval effort theories are useful in understanding the testing effect, they do not necessarily provide a mechanism through which more effortful retrieval strengthens memory.

Transfer Appropriate Processing

Transfer appropriate processing (TAP) proposes that memory is better when there is congruence between the cognitive processes enlisted during encoding and retrieval (Morris et al., 1977). TAP can be used to explain the testing effect because retrieval practice is more similar to the final test in format than restudy (or no test) and, as such, the specific processes required for successful final recall have been practiced during retrieval practice, resulting in a benefit to memory.

Evidence on whether TAP is a mechanism of the testing effect has been somewhat inconsistent. TAP theory predicts that the degree of similarity between practice and final tests should correlate with testing effect magnitude. Earlier investigations into this correlation obtained limited evidence with a fully factorial experiment crossing practice and final test format using free recall, cued recall and recognition tests finding that free recall practice tests always produced the best memory, regardless of final test format (Carpenter & DeLosh, 2006), and the meta-analysis of Rowland (2014) confirming no correlation with testing effect magnitude and format matching. In contrast, and providing some support for TAP, Veltre et al. (2015) found an influence of format matching when using tests cueing semantic or episodic processing and the meta-analysis by Yang et al. (2021) found a robust difference between matched and mismatched formats.

Theoretical Mechanisms of the Forward Testing Effect

Reduced Build-Up of Interference

One account of the forward testing effect argues that testing benefits learning by reducing the negative effect of proactive interference. This account is based on the finding that alongside potentiating new learning, testing also reduces the occurrence of previous list intrusions. For example, in Szpunar et al. (2008), prior testing not only increased the number of words recalled from the new list, it also reduced the erroneous recall of words from previous lists. Additionally, testing after each list eliminated the progressive buildup of intrusions that occurred during learning of multiple lists when the participants were not tested.

Based on these results, Szpunar et al. (2008) presented a source-monitoring framework: testing adds contextual details to the memory of each list, these details make it easier to distinguish between lists during learning, and, in turn, the search set during the criterial assessment is reduced to only the new list. This framework suggests that the ability to distinguish sources of information reduces proactive interference and results in both decreased intrusions and potentiated new learning. Evidence for this reduction in the buildup of proactive interference has also been found using different materials, such as face-name paired-associates (Weinstein et al., 2011), and confirmed in mediation analyses (Yang et al., 2022) and in studies that have found that testing reduces response latency during the criterial assessment indicating a reduced search set (Bäuml & Kliegl, 2013). Additionally, the reduction in intrusions has been confirmed by the meta-analysis by Chan, Meissner, and Davis (2018), which found that testing robustly reduced (g = -0.77) intrusions relative to comparison tasks. It is important to remember, however, that

despite a causal mechanism being proposed, the relationship between intrusions and new learning remains purely correlational and, as such, experimental manipulations of intrusions are still required to assess the account (Ahn & Chan, 2022).

Reset of Encoding Processes (ROE)

The ROE account of the forward testing effect argues that learning expends cognitive resources, mainly attention, and that this depletion negatively impacts new learning. Testing resets these resources, possibly by inducing a context change, meaning that encoding of subsequent lists is equal to that of initial lists. Restudy does not reset resources and therefore new learning is hindered. This account has been proposed by Pastötter et al. (2008), who found that a context change task, where individuals were asked to mentally travel to their parents' home, potentiated new learning and was correlated with decreased theta and alpha activation. Theta and alpha activation are thought to be indicative of diffused attention during learning, so a reduction suggests focused attention is maintained (Pastötter et al., 2011).

Mind-Wandering and New Learning

Decades of research have established that learning expends attentional resources and increases inattention and the occurrence of task irrelevant thoughts (mind-wandering), and that these, in turn, negatively impact subsequent learning (for a review, see Smallwood & Schooler, 2015). Research has found that testing may reduce the occurrence of mind-wandering, relative to restudy, and increase the number of task-relevant thoughts (e.g., Szpunar et al., 2013; Wong & Lim, 2022).

There are different accounts of why testing has this attenuating impact on mind-wandering. Task difficulty and retrieval failure during testing may recalibrate participants' perception of their own learning, indicating that greater attention during new learning is needed (Peterson & Wissman, 2020). Tests may also increase test-expectancy, the premonition that learned information will be assessed in the future (Szpunar et al., 2008), increasing motivation.

Retrieval-Induced Metacognitive Change

Another proposed mechanism of the forward testing effect is that testing induces metacognitive shifts that enhance the encoding and/or retrieval of new materials, that is, they induce a change to more efficient learning strategies. Several lines of evidence are consistent with this hypothesis. For instance, Yang et al. (2017) and Soderstrom and Bjork (2014) demonstrated that prior testing improved self-regulation of time spent during new learning. Furthermore, Chan, Manley, et al. (2018) presented evidence that prior testing may increase the usage of conceptual strategies where new materials are encoded and retrieved according to the semantic associations between them. Along the same lines, Yang et al. (2022) showed that prior testing drives participants to recall new items in a more temporally structured way.

Integration and Activation Facilitation

One final set of theories proposed to explain the forward testing effect suggests that testing potentiates new learning by facilitating the recollection or activation of originally learned materials during new learning, and the incorporation of these materials into a single coherent memory. These theories are referred to as integration theories (see Chan, Meissner, & Davis, 2018, for a review).

In one study, Wissman et al. (2011) conjectured that testing may facilitate integration as an indirect consequence of a direct benefit of retrieval. As prior tests enhance original learning (a direct effect), that consolidated knowledge is more readily accessible during new learning. When materials are related across sections, test-facilitated activation may potentiate new learning as the greater accessibility enhances comprehension. This is referred to as a mediated or indirect effect of testing and suggests that the forward testing effect may be dependent on the direct testing effect. Wissman et al. presented some evidence consistent with this account.

The benefit of facilitated activation on new learning has also been explored by Wahlheim (2015), who demonstrated that testing may reduce item-level interference through change recollection and inter-list integration. Participants learning a list of word-pairs experienced less negative impact from intruding items from previous lists when they recalled that the item was from a previous list (and not from the current list). Integration is not necessarily semantic in nature—it is possible that integration enables better source discrimination. Pierce et al. (2017) conducted an experiment in which the final assessment used a modified free recall format. Their assessment asked participants to recall items from the final list only, but also to report any items (intrusions) from a previous list that came to mind during the assessment and specify which list they were from. Pierce et al. found that the test and restudy groups reported a similar number of intrusions, but the test group was better able to specify that items were from a previous list. Enhanced source-monitoring reduced the interfering impact that intrusions had on learning. Source-monitoring was proposed by Wahlheim (2015) to explain why integration enhances learning. Change recollection during new learning allows for the creation of a single temporally structured memory trace. That is, participants recall the order in which they learn information, which reduces interference.

Conclusions and Future Directions

The burgeoning literature on test-enhanced learning comprises a range of conceptually linked phenomena that are caused by the simple act of retrieving information from memory. Although many of these phenomena have been known for decades (e.g., Spitzer, 1939), others, such as the forward testing effect and the pretesting effect, did not come under sustained attention in memory research until the 1990s (e.g., Pressley et al., 1990; Szpunar et al., 2008). One strong motivating factor behind this attention is the possibility that test-enhanced learning could have impactful translation into the classroom (Trumbo et al., 2021) and a second is that retrieval seems to offer a window through which the theoretical complexity of learning and memory can profitably be viewed.

Applied research will continue to investigate the practical application of test-enhanced learning in the classroom, but beyond the conventional classroom too, for example, in online learning (e.g., Yong & Lim, 2016), occupational training (see Green et al., 2018), and neuropsychological rehabilitation (e.g., Middleton et al., 2015). A question of the utmost importance is whether tests and quizzes can alleviate students' test anxiety. Although the research to date suggests an affirmative answer, further high-powered randomized control trials are needed. From a theoretical perspective, one important issue is whether a unifying conceptual framework can be developed to bring together some of the rather disparate accounts that have been put forward to explain various micro-phenomena. An even more fundamental challenge is to formulate computational-level explanations of the effects of retrieval on learning.

Acknowledgments

The writing of this review was supported by a grant from the U.K. Economic and Social Research Council (ES/S014616/1). We thank Philip Higham and Bernhard Pastötter for their valuable feedback on an early draft.

Further Reading

Chan, J. C. K., Meissner, C. A., & Davis, S. D. (2018). Retrieval potentiates new learning: A theoretical and meta-analytic review <https://doi.org/10.1037/bul0000166>. *Psychological Bulletin*, *144*, 1111–1146.

McDermott, K. B. (2021). Practicing retrieval facilitates learning <https://doi.org/10.1146/annurev-psych-010419-051019>. *Annual Review of Psychology*, *72*, 609–633.

Rivers, M. L. (2021). Metacognition about practice testing: A review of learners' beliefs, monitoring, and control of test-enhanced learning <https://doi.org/10.1007/s10648-020-09578-2>. *Educational Psychology Review*, *33*(3), 823–862.

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures <https://doi.org/10.1073/pnas.122176411>. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 6313–6317.

Trumbo, M. C. S., McDaniel, M. A., Hodge, G. K., Jones, A. P., Matzen, L. E., Kittinger, L. I., Kittinger, R. S., & Clark, V. P. (2021). Is the testing effect ready to be put to work? Evidence from the laboratory to the classroom <https://doi.org/10.1037/tps0000292>. *Translational Issues in Psychological Science*, *7*(3), 332–355.

Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review <https://doi.org/10.1037/bul0000309>. *Psychological Bulletin*, *147*, 399–435.

References

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing <https://doi.org/10.3102/0034654316689306>. *Review of Educational Research*, *87*, 659–701.

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 <https://doi.org/10.1016/j.jarmac.2014.07.002>. *Journal of Applied Research in Memory and Cognition*, *3*(3), 131–139.

Ahn, D., & Chan, J. C. K. (2022). Does testing enhance new learning because it insulates against proactive interference? <https://doi.org/10.3758/s13421-022-01273-7> *Memory & Cognition*, *50*, 1664–1682.

Alamri, A., & Higham, P. A. (2022). The dark side of corrective feedback: Controlled and automatic influences of retrieval practice <https://doi.org/10.1037/xlm0001138>. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(5), 752–768.

Ariel, R., & Karpicke, J. D. (2018). Improving self-regulated learning with a retrieval practice intervention <https://doi.org/10.1037/xap0000133>. *Journal of Experimental Psychology: Applied*, *24*(1), 43–56.

Bai, C.-H., Bridger, E., Zimmer, H., & Mecklinger, A. (2015). The beneficial effect of testing: An event-related potential study <https://doi.org/10.3389/fnbeh.2015.00248>. *Frontiers in Behavioral Neuroscience*, *9*, Article 248.

Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events <https://doi.org/10.3102/00346543061002213>. *Review of Educational Research*, *61*(2), 213–238.

Bäuml, K.-H. T., & Kliegl, O. (2013). The critical role of retrieval processes in release from proactive interference <https://doi.org/10.1016/j.jml.2012.07.006>. *Journal of Memory and Language*, *68*, 39–53.

Bäuml, K.-H. T., & Trißl, L. (2022). Selective memory retrieval can revive forgotten memories <https://doi.org/10.1073/pnas.2114377119>. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(8), Article e2114377119.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalf & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). MIT Press.

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (Vol. 2, pp. 35–67). Erlbaum.

Blasiman, R. N., Dunlosky, J., & Rawson, K. A. (2017). The what, how much, and when of study strategies: Comparing intended versus actual study behaviour <https://doi.org/10.1080/09658211.2016.1221974>. *Memory*, *25*, 784–792.

Bottiroli, S., Dunlosky, J., Guerini, K., Cavallini, E., & Hertzog, C. (2010). Does task affordance moderate age-related deficits in strategy production? <https://doi.org/10.1080/13825585.2010.481356> *Aging, Neuropsychology, and Cognition*, *17*(5), 591–602.

Boustani, S., Don, H. J., & Shanks, D. R. (2022, November). *Multiple short tests or one long test: Which is better for learning?* [Poster presentation]. The 63rd Annual Meeting of the Psychonomic Society, Boston.

Boustani, S., & Shanks, D. R. (2022). Heterogeneity and publication bias in research on test-potentiated new learning <https://doi.org/10.1525/collabra.31996>. *Collabra: Psychology*, *8*(1), Article 31996.

Broeren, M., Heijltjes, A., Verkoeijen, P., Smeets, G., & Arends, L. (2021). Supporting the self-regulated use of retrieval practice: A higher education classroom experiment <https://doi.org/https:/doi.org/10.1016/j.cedpsych.2020.101939>. *Contemporary Educational Psychology*, *64*, Article 101939.

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

Butler, A. C., Karpicke, J. D., & Roediger, H. L., III (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses <https://doi.org/10.1037/0278-7393.34.4.918>. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 918–928.

Butler, A. C., & Roediger, H. L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition*, *36*, 604–616.

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval <https://doi.org/10.1037/a0017021>. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 1563–1569.

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect <https://doi.org/10.1037/a0024140>. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1547–1552.

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition*, *34*, 268–276.

Carpenter, S. K., & Pashler, H. (2007). Testing beyond words: Using tests to enhance visuospatial map learning <https://doi.org/10.3758/BF03194092>. *Psychonomic Bulletin & Review*, *14*(3), 474–478.

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

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

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis <https://doi.org/10.1037/0033-2909.132.3.354>. *Psychological Bulletin*, *132*, 354–380.

Chan, J. C. K. (2009). When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing <https://doi.org/10.1016/j.jml.2009.04.004>. *Journal of Memory and Language*, *61*(2), 153–170.

Chan, J. C. K., Manley, K. D., Davis, S. D., & Szpunar, K. K. (2018). Testing potentiates new learning across a retention interval and a lag: A strategy change perspective <https://doi.org/10.1016/j.jml.2018.05.007>. *Journal of Memory and Language*, *102*, 83–96.

Chan, J. C. K., Meissner, C. A., & Davis, S. D. (2018). Retrieval potentiates new learning: A theoretical and meta-analytic review <https://doi.org/10.1037/bul0000166>. *Psychological Bulletin*, *144*, 1111–1146.

Chan, J. C. K., O'Donnell, R., & Manley, K. D. (2022). Warning weakens retrieval-enhanced suggestibility only when it is given shortly after misinformation: The critical importance of timing <https://doi.org/10.1037/xap0000394>. *Journal of Experimental Psychology: Applied*.

Chan, J. C. K., Thomas, A. K., & Bulevich, J. B. (2009). Recalling a witnessed event increases eyewitness suggestibility: The reversed testing effect <https://doi.org/10.1111/j.1467-9280.2008.02245.x>. *Psychological Science*, *20*(1), 66–73.

Cho, K. W., Neely, J. H., Brennan, M. K., Vitrano, D., & Crocco, S. (2017). Does testing increase spontaneous mediation in learning semantically related paired associates? <https://doi.org/10.1037/xlm0000414> *Journal of Experimental Psychology: Learning*, *Memory, and Cognition*, *43*, 1768–1778.

Coppens, L. C., Verkoeijen, P. P. J. L., & Rikers, R. M. J. P. (2011). Learning Adinkra symbols: The effect of testing <https://doi.org/10.1080/20445911.2011.507188>. *Journal of Cognitive Psychology*, *23*(3), 351–357.

Davis, S. D., Chan, J. C. K., & Wilford, M. M. (2017). The dark side of interpolated testing: Frequent switching between retrieval and encoding impairs new learning <https://doi.org/10.1016/j.jarmac.2017.07.002>. *Journal of Applied Research in Memory and Cognition*, *6*, 434–441.

Don, H. J., Yang, C., Boustani, S., & Shanks, D. R. (2022). Do partial and distributed tests enhance new learning? <https://doi.org/10.1037/xap0000440> *Journal of Experimental Psychology: Applied*.

Endres, T., Carpenter, S., Martin, A., & Renkl, A. (2017). Enhancing learning by retrieval: Enriching free recall with elaborative prompting <https://doi.org/https:/doi.org/10.1016/j.learninstruc.2016.11.010>. *Learning and Instruction*, *49*, 13–20.

Erdman, M. R., & Chan, J. C. K. (2013). Providing corrective feedback during retrieval practice does not increase retrieval-induced forgetting <https://doi.org/10.1080/20445911.2013.790389>. *Journal of Cognitive Psychology*, *25*(6), 692–703.

Finley, J. R., Tullis, J. G., & Benjamin, A. S. (2010). Metacognitive control of learning and remembering <https://doi.org/10.1007/978-1-4419-5716-0_6>. In M. S. Khine & I. M. Saleh (Eds.), *New science of learning: Cognition, computers and collaboration in education* (pp. 109–131). Springer.

Finn, B., & Roediger, H. L. (2013). Interfering effects of retrieval in learning new information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1665–1681.

Gates, A. I. (1917). Recitation as a factor in memorizing. *Archives of Psychology*, *40*, 1–104.

Green, M. L., Moeller, J. J., & Spak, J. M. (2018). Test-enhanced learning in health professions education: A systematic review: BEME Guide No. 48 <https://doi.org/10.1080/0142159X.2018.1430354>. *Medical Teacher*, *40*(4), 337–350.

Hattie, J., & Timperley, H. (2007). The power of feedback <https://doi.org/10.3102/003465430298487>. *Review of Educational Research*, *77*(1), 81–112.

Healy, A. F., Jones, M., Lalchandani, L. A., & Tack, L. A. (2017). Timing of quizzes during learning: Effects on motivation and retention <https://doi.org/10.1037/xap0000123>. *Journal of Experimental Psychology: Applied*, *23*, 128–137.

Heitmann, S., Grund, A., Berthold, K., Fries, S., & Roelle, J. (2018). Testing is more desirable when it is adaptive and still desirable when compared to note-taking <https://doi.org/10.3389/fpsyg.2018.02596>. *Frontiers in Psychology*, *9*, Article 2596.

Higham, P. A., Zengel, B., Bartlett, L. K., & Hadwin, J. A. (2022). The benefits of successive relearning on multiple learning outcomes <https://doi.org/10.1037/edu0000693>. *Journal of Educational Psychology*, *114*, 928–944.

Hui, L., de Bruin, A. B. H., Donkers, J., & van Merriënboer, J. J. G. (2021). Does individual performance feedback increase the use of retrieval practice? <https://doi.org/10.1007/s10648-021-09604-x> *Educational Psychology Review*, *33*(4), 1835–1857.

Jacoby, L. L., Wahlheim, C. N., & Coane, J. H. (2010). Test-enhanced learning of natural concepts: Effects on recognition memory, classification, and metacognition <https://doi.org/10.1037/a0020636>. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1441–1451.

Jing, H. G., Szpunar, K. K., & Schacter, D. L. (2016). Interpolated testing influences focused attention and improves integration of information during a video-recorded lecture <https://doi.org/10.1037/xap0000087>. *Journal of Experimental Psychology: Applied*, *22*, 305–318.

Kang, S. H. K. (2010). Enhancing visuospatial learning: The benefit of retrieval practice <https://doi.org/10.3758/MC.38.8.1009>. *Memory & Cognition*, *38*(8), 1009–1017.

Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning <https://doi.org/10.1037/a0017341>. *Journal of Experimental Psychology: General*, *138*, 469–486.

Karpicke, J. D. (2017). Retrieval-based learning: A decade of progress <https://doi.org/10.1016/B978-0-12-809324-5.21055-9>. In J. Wixted (Ed.), *Learning and memory: A comprehensive reference: Vol. 2: Cognitive psychology of memory* (pp. 487–514). Academic Press.

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping <https://doi.org/10.1126/science.1199327>. *Science*, *331*(6018), 772–775.

Karpicke, J. D., Blunt, J. R., Smith, M. A., & Karpicke, S. S. (2014). Retrieval-based learning: The need for guided retrieval in elementary school children <https://doi.org/10.1037/h0101802>. *Journal of Applied Research in Memory and Cognition*, *3*(3), 198–206.

Karpicke, J. D., Butler, A. C., & Roediger, H. L., III. (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? <https://doi.org/10.1080/09658210802647009> *Memory*, *17*(4), 471–479.

Kay, A. E., Hardy, J., & Galloway, R. K. (2020). Student use of PeerWise: A multi-institutional, multidisciplinary evaluation <https://doi.org/10.1111/bjet.12754>. *British Journal of Educational Technology*, *51*, 25–35.

Khanna, M. M. (2015). Ungraded pop quizzes: Test-enhanced learning without all the anxiety <https://doi.org/10.1177/0098628315573144>. *Teaching of Psychology*, *42*(2), 174–178.

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

Kromann, C. B., Bohnstedt, C., Jensen, M. L., & Ringsted, C. (2010). The testing effect on skills learning might last 6 months <https://doi.org/10.1007/s10459-009-9207-x>. *Advances in Health Sciences Education*, *15*(3), 395–401.

Kuhbandner, C., & Emmerdinger, K. J. (2019). Do students really prefer repeated rereading over testing when studying textbooks? A reexamination <https://doi.org/10.1080/09658211.2019.1610177>. *Memory*, *27*(7), 952–961.

Kulhavy, R. W. (1977). Feedback in written instruction <https://doi.org/10.3102/00346543047002211>. *Review of Educational Research*, *47*(2), 211–232.

Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude <https://doi.org/10.1007/BF01320096>. *Educational Psychology Review*, *1*(4), 279–308.

Kulik, J. A., & Kulik, C.-L. C. (1988). Timing of feedback and verbal learning <https://doi.org/10.3102/00346543058001079>. *Review of Educational Research*, *58*(1), 79–97.

Latimier, A., Riegert, A., Ly, S. T., & Ramus, F. (2020). Retrieval practice promotes long-term retention irrespective of the placement <https://psyarxiv.com/dk63q/>. *PsyArXiv*.

Latimier, A., Riegert, A., Peyre, H., Ly, S. T., Casati, R., & Ramus, F. (2019). Does pre-testing promote better retention than post-testing? <https://doi.org/10.1038/s41539-019-0053-1> *NPJ Science of Learning*, *4*, 15, Article 15.

Lavigne, E., & Risko, E. F. (2018). Optimizing the use of interpolated tests: The influence of interpolated test lag <https://doi.org/10.1037/stl0000118>. *Scholarship of Teaching and Learning in Psychology*, *4*(4), 211–221.

Lee, H. S., & Ahn, D. (2018). Testing prepares students to learn better: The forward effect of testing in category learning <https://doi.org/10.1037/edu0000211>. *Journal of Educational Psychology*, *110*, 203–217.

Lefevre, D., & Cox, B. (2017). Delayed instructional feedback may be more effective, but is this contrary to learners' preferences? <https://doi.org/https:/doi.org/10.1111/bjet.12495> *British Journal of Educational Technology*, *48*(6), 1357–1367.

Lehman, M., & Karpicke, J. D. (2016). Elaborative retrieval: Do semantic mediators improve memory? <https://doi.org/10.1037/xlm0000267> *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 1573–1591.

Lehman, M., Smith, M. A., & Karpicke, J. D. (2014). Toward an episodic context account of retrieval-based learning: Dissociating retrieval practice and elaboration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 1787–1794.

Little, J. L. (2018). The role of multiple-choice tests in increasing access to difficult-to-retrieve information <https://doi.org/10.1080/20445911.2018.1492581>. *Journal of Cognitive Psychology*, *30*(5–6), 520–531.

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.

Lortie-Forgues, H., & Inglis, M. (2019). Rigorous large-scale educational RCTs are often uninformative: Should we be concerned? <https://doi.org/10.3102/0013189X19832850> *Educational Researcher*, *48*, 158–166.

McAndrew, M., Morrow, C. S., Atiyeh, L., & Pierre, G. C. (2016). Dental student study strategies: Are self-testing and scheduling related to academic performance? <https://doi.org/10.1002/j.0022-0337.2016.80.5.tb06114.x> *Journal of Dental Education*, *80*(5), 542.

McCabe, J. (2011). Metacognitive awareness of learning strategies in undergraduates <https://doi.org/10.3758/s13421-010-0035-2>. *Memory & Cognition*, *39*, 462–476.

McCabe, J. A., & Lummis, S. N. (2018). Why and how do undergraduates study in groups? <https://doi.org/http:/dx.doi.org/10.1037/stl0000099> *Scholarship of Teaching and Learning in Psychology*, *4*(1), 27–42.

McDermott, K. B. (2021). Practicing retrieval facilitates learning <https://doi.org/10.1146/annurev-psych-010419-051019>. *Annual Review of Psychology*, *72*, 609–633.

Middleton, E. L., Schwartz, M. F., Rawson, K. A., & Garvey, K. (2015). Test-enhanced learning versus errorless learning in aphasia rehabilitation: Testing competing psychological principles <https://doi.org/10.1037/xlm0000091>. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(4), 1253–1261.

Morehead, K., Rhodes, M. G., & DeLozier, S. (2016). Instructor and student knowledge of study strategies <https://doi.org/10.1080/09658211.2014.1001992>. *Memory*, *24*(2), 257–271.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 519–533.

Mullet, H. G., Butler, A. C., Verdin, B., von Borries, R., & Marsh, E. J. (2014). Delaying feedback promotes transfer of knowledge despite student preferences to receive feedback immediately <https://doi.org/https:/doi.org/10.1016/j.jarmac.2014.05.001>. *Journal of Applied Research in Memory and Cognition*, *3*(3), 222–229.

Mulligan, N. W. (2020). Negative effects of repetition and testing. In A. M. Cleary & B. L. Schwartz (Eds.), *Memory quirks: The study of odd phenomena in memory*. Routledge.

Murayama, K., Miyatsu, T., Buchli, D., & Storm, B. C. (2014). Forgetting as a consequence of retrieval: A meta-analytic review of retrieval-induced forgetting <https://doi.org/10.1037/a0037505>. *Psychological Bulletin*, *140*(5), 1383–1409.

Pan, S. C., & Rickard, T. C. (2018). Transfer of test-enhanced learning: Meta-analytic review and synthesis <https://doi.org/10.1037/bul0000151>. *Psychological Bulletin*, *144*(7), 710–756.

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? <https://doi.org/10.1037/0278-7393.31.1.3> *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 3–8.

Pastötter, B., & Bäuml, K.-H. T. (2014). Retrieval practice enhances new learning: The forward effect of testing <https://doi.org/10.3389/fpsyg.2014.00286>. *Frontiers in Psychology*, *5*, Article 286.

Pastötter, B., & Bäuml, K.-H. T. (2016). Reversing the testing effect by feedback: Behavioral and electrophysiological evidence. *Cognitive, Affective, & Behavioral Neuroscience*, *16*, 473–488.

Pastötter, B., Bäuml, K.-H. T., & Hanslmayr, S. (2008). Oscillatory brain activity before and after an internal context change: Evidence for a reset of encoding processes <https://doi.org/https:/doi.org/10.1016/j.neuroimage.2008.07.005>. *NeuroImage*, *43*(1), 173–181.

Pastötter, B., Schicker, S., Niedernhuber, J., & Bäuml, K.-H. T. (2011). Retrieval during learning facilitates subsequent memory encoding <https://doi.org/10.1037/a0021801>. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 287–297.

Peeck, J. (1970). Effect of prequestions on delayed retention of prose material <https://doi.org/10.1037/h0029104>. *Journal of Educational Psychology*, *61*(3), 241–246.

Persky, A. M. (2018). A four year longitudinal study of student learning strategies <https://doi.org/https:/doi.org/10.1016/j.cptl.2018.08.012>. *Currents in Pharmacy Teaching and Learning*, *10*(11), 1496–1500.

Peterson, D. J., & Wissman, K. (2020). Using tests to reduce mind-wandering during learning review <https://doi.org/10.1080/09658211.2020.1748657>. *Memory*, *28*(4), 582–587.

Pierce, B. H., Gallo, D. A., & McCain, J. L. (2017). Reduced interference from memory testing: A postretrieval monitoring account <https://doi.org/10.1037/xlm0000377>. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(7), 1063–1072.

Pressley, M., Tanenbaum, R., McDaniel, M. A., & Wood, E. (1990). What happens when university students try to answer prequestions that accompany textbook material? <https://doi.org/https:/doi.org/10.1016/0361-476X(90)90003-J> *Contemporary Educational Psychology*, *15*(1), 27–35.

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? <https://doi.org/10.1016/j.jml.2009.01.004> *Journal of Memory and Language*, *60*(4), 437–447.

Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science*, *330*, 335.

Racsmány, M., Szőllősi, Á., & Marián, M. (2020). Reversing the testing effect by feedback is a matter of performance criterion at practice <https://doi.org/10.3758/s13421-020-01041-5>. *Memory & Cognition*, *48*(7), 1161–1170.

Rawson, K. A., Vaughn, K. E., & Carpenter, S. K. (2015). Does the benefit of testing depend on lag, and if so, why? Evaluating the elaborative retrieval hypothesis <https://doi.org/10.3758/s13421-014-0477-z>. *Memory & Cognition*, *43*(4), 619–633.

Rivers, M. L. (2021). Metacognition about practice testing: A review of learners' beliefs, monitoring, and control of test-enhanced learning <https://doi.org/10.1007/s10648-020-09578-2>. *Educational Psychology Review*, *33*(3), 823–862.

Roediger, H. L., Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied*, *17*, 382–395.

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

Roediger, H. L., & Karpicke, J. D. (2006b). The power of testing memory: Basic research and implications for educational practice <https://doi.org/10.1111/j.1745-6916.2006.00012.x>. *Perspectives on Psychological Science*, *1*, 181–210.

Roediger, H. L., & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing <https://doi.org/10.1037/0278-7393.31.5.1155>. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 1155–1159.

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect <https://doi.org/10.1037/a0037559>. *Psychological Bulletin*, *140*, 1432–1463.

Rowland, C. A., & DeLosh, E. L. (2015). Mnemonic benefits of retrieval practice at short retention intervals <https://doi.org/10.1080/09658211.2014.889710>. *Memory*, *23*(3), 403–419.

Sagaria, S. D., & Di Vesta, F. J. (1978). Learner expectations induced by adjunct questions and the retrieval of intentional and incidental information <https://doi.org/10.1037/0022-0663.70.3.280>. *Journal of Educational Psychology*, *70*(3), 280–288.

Schwieren, J., Barenberg, J., & Dutke, S. (2017). The testing effect in the psychology classroom: A meta-analytic perspective <https://doi.org/10.1177/1475725717695149>. *Psychology Learning & Teaching*, *16*, 179–196.

Schwoebel, J., Depperman, A. K., & Scott, J. L. (2018). Distinct episodic contexts enhance retrieval-based learning <https://doi.org/10.1080/09658211.2018.1464190>. *Memory*, *26*(9), 1291–1296.

Sennhenn-Kirchner, S., Goerlich, Y., Kirchner, B., Notbohm, M., Schiekirka, S., Simmenroth, A., & Raupach, T. (2018). The effect of repeated testing vs repeated practice on skills learning in undergraduate dental education <https://doi.org/10.1111/eje.12254>. *European Journal of Dental Education*, *22*(1), e42–e47.

Smallwood, J., & Schooler, J. W. (2015). The science of mind wandering: Empirically navigating the stream of consciousness <https://doi.org/10.1146/annurev-psych-010814-015331>. *Annual Review of Psychology*, *66*, 487–518.

Soderstrom, N. C., & Bjork, R. A. (2014). Testing facilitates the regulation of subsequent study time <https://doi.org/10.1016/j.jml.2014.03.003>. *Journal of Memory and Language*, *73*, 99–115.

Spitzer, H. F. (1939). Studies in retention <https://doi.org/10.1037/h0063404>. *Journal of Educational Psychology*, *30*, 641–656.

Storm, B. C., Friedman, M. C., Murayama, K., & Bjork, R. A. (2014). On the transfer of prior tests or study events to subsequent study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 115–124.

Storm, B. C., & Soares, J. S. (2022). Relearning can eliminate the effect of retrieval-induced forgetting <https://doi.org/10.1007/s00426-021-01601-7>. *Psychological Research*, *86*(6), 1725–1736.

Sullivan, D. (2017). Mediating test anxiety through the testing effect in asynchronous, objective, online assessments at the university level <https://doi.org/10.5296/jet.v4i2.10777>. *Journal of Education and Training*, *4*(2), 107–123.

Szpunar, K. K., Jing, H. G., & Schacter, D. L. (2014). Overcoming overconfidence in learning from video-recorded lectures: Implications of interpolated testing for online education <https://doi.org/10.1016/j.jarmac.2014.02.001>. *Journal of Applied Research in Memory and Cognition*, *3*, 161–164.

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures <https://doi.org/10.1073/pnas.122176411>. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 6313–6317.

Szpunar, K. K., McDermott, K. B., & Roediger, H. L. (2008). Testing during study insulates against the buildup of proactive interference <https://doi.org/10.1037/a0013082>. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1392–1399.

Toppino, T. C., LaVan, M. H., & Iaconelli, R. T. (2018). Metacognitive control in self-regulated learning: Conditions affecting the choice of restudying versus retrieval practice <https://doi.org/10.3758/s13421-018-0828-2>. *Memory & Cognition*, *46*, 1164–1177.

Trumbo, M. C. S., McDaniel, M. A., Hodge, G. K., Jones, A. P., Matzen, L. E., Kittinger, L. I., Kittinger, R. S., & Clark, V. P. (2021). Is the testing effect ready to be put to work? Evidence from the laboratory to the classroom <https://doi.org/10.1037/tps0000292>. *Translational Issues in Psychological Science*, *7*(3), 332–355.

Tullis, J. G., Fiechter, J. L., & Benjamin, A. S. (2018). The efficacy of learners' testing choices <https://doi.org/10.1037/xlm0000473>. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 540–552.

Tullis, J. G., Finley, J. R., & Benjamin, A. S. (2013). Metacognition of the testing effect: Guiding learners to predict the benefits of retrieval <https://doi.org/10.3758/s13421-012-0274-5>. *Memory & Cognition*, *41*(3), 429–442.

Tulving, E., & Watkins, M. J. (1974). Negative transfer: Effects of testing one list on recall of another <https://doi.org/10.1016/s0022-5371(74)80043-5>. *Journal of Verbal Learning and Verbal Behavior*, *13*, 181–193.

Uner, O., & Roediger, H. L. (2018). The effect of question placement on learning from textbook chapters <https://doi.org/10.1016/j.jarmac.2017.09.002>. *Journal of Applied Research in Memory and Cognition*, *7*, 116–122.

Vaughn, K. E., & Kornell, N. (2019). How to activate students' natural desire to test themselves <https://doi.org/10.1186/s41235-019-0187-y>. *Cognitive Research: Principles and Implications*, *4*, Article 35.

Veltre, M. T., Cho, K. W., & Neely, J. H. (2015). Transfer-appropriate processing in the testing effect <https://doi.org/10.1080/09658211.2014.970196>. *Memory*, *23*, 1229–1237.

Wahlheim, C. N. (2015). Testing can counteract proactive interference by integrating competing information <https://doi.org/10.3758/s13421-014-0455-5>. *Memory & Cognition*, *43*(1), 27–38.

Weinstein, Y., McDermott, K. B., & Szpunar, K. K. (2011). Testing protects against proactive interference in face-name learning <https://doi.org/10.3758/s13423-011-0085-x>. *Psychonomic Bulletin & Review*, *18*, 518–523.

Weinstein, Y., Nunes, L. D., & Karpicke, J. D. (2016). On the placement of practice questions during study <https://doi.org/10.1037/xap0000071>. *Journal of Experimental Psychology: Applied*, *22*, 72–84.

Whiffen, J. W., & Karpicke, J. D. (2017). The role of episodic context in retrieval practice effects <https://doi.org/10.1037/xlm0000379>. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*, 1036–1046.

Wissman, K. T., & Rawson, K. A. (2015). Grain size of recall practice for lengthy text material: Fragile and mysterious effects on memory <https://doi.org/10.1037/xlm0000047>. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(2), 439–455.

Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2011). The interim test effect: Testing prior material can facilitate the learning of new material <https://doi.org/10.3758/s13423-011-0140-7>. *Psychonomic Bulletin & Review*, *18*, 1140–1147.

Wong, S. S. H., & Lim, S. W. H. (2022). A mind-wandering account of the testing effect: Does context variation matter? <https://doi.org/10.3758/s13423-021-01989-8> *Psychonomic Bulletin & Review*, *29*(1), 220–229.

Yang, C., Chew, S.-J., Sun, B. K., & Shanks, D. R. (2019). The forward effects of testing transfer to different domains of learning <https://doi.org/10.1037/edu0000320>. *Journal of Educational Psychology*, *111*(5), 809–826.

Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review <https://doi.org/10.1037/bul0000309>. *Psychological Bulletin*, *147*, 399–435.

Yang, C., Potts, R., & Shanks, D. R. (2018). Enhancing learning and retrieval of new information: A review of the forward testing effect <https://doi.org/10.1038/s41539-018-0024-y>. *NPJ Science of Learning*, *3*(1), Article 8.

Yang, C., Zhao, W., Li, J., Luo, L., & Shanks, D. R. (2023). *Do practice tests (quizzes) reduce or provoke test anxiety? A meta-analytic review*. Unpublished manuscript.

Yang, C., Zhao, W., Luo, L., Sun, B., Potts, R., & Shanks, D. R. (2022). Testing potential mechanisms underlying test-potentiated new learning <https://doi.org/10.1037/xlm0001021>. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(8), 1127-1143.

Yang, C. L., Potts, R., & Shanks, D. R. (2017). The forward testing effect on self-regulated study time allocation and metamemory monitoring <https://doi.org/10.1037/xap0000122>. *Journal of Experimental Psychology: Applied*, *23*(3), 263–277.

Yang, C. L., & Shanks, D. R. (2018). The forward testing effect: Interim testing enhances inductive learning <https://doi.org/10.1037/xlm0000449>. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(3), 485–492.

Yang, C. L., Sun, B. K., Potts, R., Yu, R. J., Luo, L., & Shanks, D. R. (2020). Do working memory capacity and test anxiety modulate the beneficial effects of testing on new learning? <https://doi.org/10.1037/xap0000278> *Journal of Experimental Psychology: Applied*, *26*(4), 724–738.

Yong, P. Z., & Lim, S. W. H. (2016). Observing the testing effect using *Coursera* video-recorded lectures: A preliminary study <https://doi.org/10.3389/fpsyg.2015.02064>. *Frontiers in Psychology*, *6*, Article 2064.

Related Articles

Dyslexia

Culture and Intelligence

Early Cognitive Development: Five Lessons from Infant Learning