

# The Value of Standardized Testing: A Perspective From Cognitive Psychology

Policy Insights from the  
Behavioral and Brain Sciences  
2015, Vol. 2(1) 13–23  
© The Author(s) 2015  
DOI: 10.1177/2372732215601116  
bbs.sagepub.com



Aaron S. Benjamin<sup>1</sup> and Hal Pashler<sup>2</sup>

## Abstract

Recent years have seen an increased push toward the standardization of education in the United States. At the federal level, both major national political parties have generally supported the institution of national guidelines known as *Common Core*—a curriculum developed by states and by philanthropic organizations. A key component of past and present educational reform measures has been standardization of tests. However, increased reliance upon tests has elicited criticism, limiting their popular acceptance and widespread adoption. Tests are not only useful for assessment purposes, however. The goal of this article is to review evidence from the recent literature in psychology that indicates that tests produce direct educational benefits for students. A reconsideration of how and how many tests are implemented based on these principles may help soften the focus on testing solely as a means of assessment and help promote wider recognition of the role of tests as potent instructional interventions.

## Keywords

learning, testing, memory, standardized testing, Common Core, education, education reform

## Tweet

Testing benefits learning. Can Common Core take advantage?

## Key Points

- Educational reform will include a means of standardizing testing.
- Testing is useful for more than just assessment.
- Tests can promote learning directly.
- The development of standardized tests should take into account the ways in which tests are known to benefit learning.

Educational outcomes in the United States are a source of concern. In a recent comprehensive assessment (Pearson, 2014), the United States ranked 14th on a composite measure of cognitive skills and educational attainment, behind countries such as Russia and Poland. Every country ranked ahead of the United States has a lower per-capita gross domestic product (GDP; World Economic Outlook Database, 2014), a fact that points strongly toward deficiencies in American education policy.

The regularization of educational standards is one means by which such poor educational outcomes in the United States are being addressed. Nations achieving top rankings in educational attainment typically have greater oversight of standards, curriculum, and testing than the United States. In

South Korea, which ranked first on the Pearson assessment, a single national ministry oversees the national curriculum and revises it regularly. In the United States, the development of a national curriculum is fraught with political consequences due to widespread concerns about federal intrusion upon local authority. For these reasons and others, the Common Core delineates general standards for achievement but carefully avoids recommending specific programs or materials for achieving those goals.

Assessing whether or not students are attaining the new criteria calls for standardization of testing. The *Race to the Top* program of the Obama administration provided substantial grants in 2010 for the development of such tests; at one point, almost all of the states had joined one or both of the major consortia (the *Partnership for Assessment of Readiness for College and Careers*, or PARCC and the *Smarter Balanced Assessment Consortium*, or SBAC) that are developing and implementing the tests. Yet those tests have proven unpopular, both on the political and the parental front. By June 2015, more than half of the states that originally joined

<sup>1</sup>University of Illinois at Urbana–Champaign, USA

<sup>2</sup>University of California, San Diego, USA

## Corresponding Author:

Aaron S. Benjamin, University of Illinois at Urbana–Champaign, 603 E. Daniel Street, Champaign, IL 61820, USA.  
Email: asbenjam@illinois.edu

have now abandoned the consortia (Wurman, 2015), and there is an unusually diverse coalition of opponents from across the political spectrum (Harris, 2015). Some of the negative feeling certainly reflects ongoing widespread misperceptions about the purview of Common Core (Clement & Brown, 2015), which only covers math and reading, and about changes to favored local curricula in response to Common Core standards. But some of the public anger is directed specifically at the tests. The development and scoring of tests is a difficult endeavor, and there is no doubt that it will take time to deliver a well calibrated and fair test in a technologically seamless way to millions of students yearly. However, there can be no doubt that a necessary component of educational reform is the adoption of educational standards, as well as a means for assessing their attainment.

We will argue here that an important but overlooked part of the conversation about how those standards and the tests should be developed is the question of what happens to the student who is taking a test. Tests, we will show, provide an opportunity to *enhance* as well as measure American students' education. We argue that a greater understanding of the benefits of taking tests can serve to defuse some of the concern over standardized testing and enable more concrete progress toward the shared goal of improving American education outcomes.

## The Historical Role of Testing in Psychology

Testing—or, more accurately, the use of tests as assessments—has a long history of research in psychology and of profitable importation into applied domains. Francis Galton, the eminent English biologist, developed what were probably the first mental tests in service of characterizing the resemblance between biological relatives (Galton, 1879). His development of rating scale and questionnaire methods led to the development of impartial assessment of traits and abilities, a point that was critical for growth of psychology as a separate discipline. Shortly after Galton's work was published, James McKeen Cattell (1890) noted that the discipline of psychology could never “attain the certainty and exactness of the physical sciences without systematic reliance on the objective measurement of human abilities” (p. 373). Tests developed by Cattell included recognizable precursors of modern cognitive and intelligence testing, including memory capacity and reaction time.

The industry of testing grew rapidly during the 20th century, largely in response to the need to develop rapid screening tests for Army recruits. *Army Alpha* and *Army Beta*, developed for use in World War I, were the first of countless standardized tests designed with many of the same restrictions and goals in mind as are currently being sought for general education standards: an emphasis on *reliability*—a person taking the test twice, or taking two different versions

of the test, should not score markedly different across those occasions—and *validity*—the test should predict what you have designed it to predict. Standardization is the key to both of these qualities.

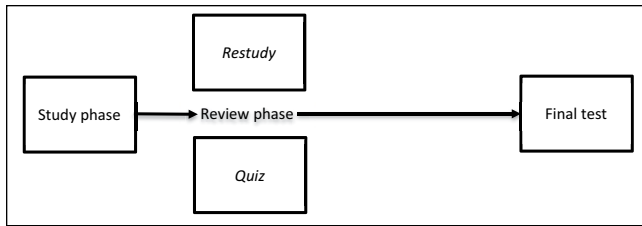
Until recently, a great deal of the most influential psychological theory on testing began and ended with these very questions, providing guidance on how to develop, implement, and score tests possessing these desirable characteristics. More recently, primarily in the last few decades, evidence has emerged out of memory research within cognitive psychology, a different historical tradition that has placed little focus on individual differences and more on mental processes common to all. This work indicates that tests have value not previously considered by test designers. The purpose of this article is to briefly review this body of research. The key point we will argue for is that the benefits of testing are not limited to those arising from good assessment: There is an important potential role for tests as *tools for learning* and not just *tools for assessment*. Following our review, we revisit the question of standardized testing and provide some recommendations for ways in which the beneficial consequences of testing can be maximized without compromising their use for assessment purposes.

## The Cognitive Benefits of Testing

A widely held view of testing essentially likens a good test to a mirror. When held up to a student, it faithfully reflects his or her knowledge and skills back to the test administrator. It is true that testing reveals what we do and do not know, with some limitations. But unlike a mirror, it also *changes what we know*. It affects our ongoing and future learning. It changes our focus of attention and can redirect our study efforts. All of these consequences have been shown to influence learning, memory, and inference, mostly in positive ways. Here we review, with some examples from cognitive psychology, some of the ways in which these beneficial changes take place.

## How Psychologists Study the Effect of Testing

To describe how psychologists have explored the effects of testing on learning, we need a bit of terminology. Figure 1 shows a design widely used in studies performed in this area. Learners usually start by learning some content, which can range from simple word lists or pictures to more educationally relevant materials such as passages from textbooks or educational videos. After this study phase, there is a review phase in which learners are asked to either re-study the material or take a test on the material. Sometimes each learner will have both types of reviews (but for different materials), and sometimes different learners will have the two different types of reviews. Later, on a final test, often after a considerable delay, learners are tested on the material (and sometimes



**Figure 1.** The typical experimental procedure by which the effects of testing are assessed.

on additional material as well) and the effects of the review phase are assessed.

To avoid confusion, whenever we refer specifically to a test during the review phase of an experiment, we will refer to this test as a *quiz*. This term should not be taken to imply anything about the nature of this test/quiz. It is only used to more clearly differentiate the event during the review phase from the final test. When we refer to testing in the generic sense, rather than to its role in a specific experiment, we will use the term *test*.

### Tests Improve Memory

When we take a test on which we are asked to retrieve and produce previously learned information, successfully recalling that information increases our ability to retrieve it again later. A good example of the advantages of testing is provided by Roediger and Karpicke (2006; Experiment 2). In their experiment, subjects read text passages, and then were either given three opportunities to re-read the passage, two additional re-reading opportunities followed by a quiz in which they tried to recall as much as they could from the passage, or three quizzes with no re-study opportunities. On a final test 1 week later, the latter group remembered the material best—despite having had only one opportunity to read the passage! Quizzing during a review phase has been shown to improve memory for other types of materials as well, including foreign-language vocabulary (Carrier & Pashler, 1992) and simple facts (McDaniel & Fisher, 1991).

Testing also increases the effectiveness of the way in which we choose to access and organize the tested information. For example, when studying a list of categorized materials, quizzes increase both the number of categories that are reported on a final test and the number of items from each of those categories (Zaromb & Roediger, 2010). These beneficial effects are probably due to the fact that testing promotes clustering of similar items during the test, a retrieval strategy that is very effective (Mulligan, 2005).

Other research has found beneficial effects of testing for a variety of different testing formats. Taking either a short-answer or a multiple-choice practice quiz enhances memory on a later test, even when the later test is in a different format than the quiz. The benefits are especially prominent if the

quiz includes feedback (LaPorte & Voss, 1975). However, it does appear overall that short-answer quizzes increase later retention to a greater degree than multiple-choice quizzes (cf. Glover, 1989; Kang, McDermott, & Roediger, 2007).

These results generalize to actual classroom settings. Students who take periodic tests on material remember that material better on later exams, and the enhancement is greater for short-answer than multiple-choice tests. This result has been shown in college students (McDaniel, Anderson, Derbish, & Morrisette, 2007), sixth-grade students (Roediger, Agarwal, McDaniel, & McDermott, 2011), and eighth-grade students in science (McDaniel, Agarwal, Huelser, McDermott, & Roediger, 2011) and history (Carpenter, Pashler, & Cepeda, 2009). It is evident for a wide range of materials, including biology (McDaniel et al., 2011), social studies (Roediger et al., 2011), general science (Roediger & Karpicke, 2006), biographical materials (Gates, 1917), and spelling (Forlano, 1936).

Are the benefits of testing simply a consequence of changes in motivation or desire to learn? Some authors have suggested that asking a question can enhance a learner's curiosity to know the answer (Berlyne, 1966). The benefits of quizzing are roughly the same regardless of how much learners are paid for their correct responses (Kang & Pashler, 2014). These results are inconsistent with the idea that motivation plays a major role in producing the benefits of testing. However, motivation does play a major role in how people direct their study and other more indirect ways in which the experience of taking tests can influence memory. These are important points we will return to in greater detail later in this article.

When taken together, these results help us understand why students who take more tests in the classroom tend to perform better on later exams (Bangert-Drowns, Kulik, & Kulik, 1991). Most of the benefits come from the first few tests, indicating that it does not require much compromise in the allocation of class time to administer periodic tests. In addition, students of all abilities appear to benefit from the opportunity to take tests (Pan, Pashler, Potter, & Rickard, 2015). As we will see below, these benefits are not limited to enhanced memory for the tested material. We will review additional research that indicates a positive role for testing for other, untested material, as well as for other aspects of cognition and motivation.

### Tests Reduce Forgetting

The cognitive benefits of testing are not like a single shot in the arm. Taking a test improves memory for the material, and it also *decreases* the rate at which we forget that material. What this means is that the benefits of testing are even greater when looking at longer-term retention. In the study with texts reviewed earlier (Roediger & Karpicke, 2006), the benefits of multiple quizzes were largest at a 1-week delay after the original study event.

All of this is particularly noteworthy because, counterintuitively, there are not many cognitive interventions that appear to slow the rate of forgetting. Studying material more leads to a higher initial degree of learning but does not slow forgetting (Anderson & Schooler, 1991; Hellyer, 1962). Employing a “deep” level of processing—in which the learner is encouraged to think about the meaning of the to-be-learned information—does not slow forgetting (Nelson & Vining, 1978). Yet, testing slows forgetting (Carpenter, Pashler, Wixted, & Vul, 2008), sometimes considerably (Wheeler, Ewers, & Buonanno, 2003), which may make it an ideal technique for promoting long-term, durable learning.

It may even be the case that the benefit to memory from testing is due *entirely* to the reduced forgetting it engenders. When a test is administered immediately after the review phase in an experiment, performance is often superior following a re-study than a quiz event. However, this advantage is short lived: after a relatively short time interval, the benefits of quizzing are apparent. So, although quizzing may not be the study regimen of choice for a student who is doing last-minute cramming, it is a better way to promote long-term retention.

*Effective organization of a sequence of tests.* The effective organization of a series of tests on the same material can enhance the benefits of testing yet further. The fact that testing decreases the rate of forgetting can be leveraged to start thinking about how tests can be efficiently sequenced. Because the material will be forgotten a little more slowly after each test, then if all tests were equally difficult from an objective standpoint, each test would actually be subjectively a little easier than the last. To render each test more similar in difficulty from the test taker’s perspective requires each test to be a little more objectively difficult than the last.

One way in which this can be done is by using an *expanding* test schedule, in which each quiz is administered at a slightly longer interval than the last one. Expanding schedules have been shown to enhance memory for names (Landauer & Bjork, 1978) and text (Storm, Bjork, & Storm, 2010). It has been used to aid learning in young children (Fritz, Morris, Nolan, & Singleton, 2007), memory-impaired populations (Camp, 2006; Schacter, Rich, & Stamp, 1985), and even in rehabilitative regimens (Wilson, Baddeley, Evans, & Sheil, 1994). They may be particularly useful for maintaining high levels of retention over long periods (Kang, Lindsey, Mozer, & Pashler, 2014). However, care must be taken to ensure that the spacing of the tests corresponds at least roughly to the rate of forgetting; if a single test is too difficult, material that is not successfully remembered on that test is unlikely to be recovered on future tests or on the final test. This scheduling difficulty may underlie cases in which the benefits of an expanding schedule are not evident when compared with evenly spaced quizzes (Logan & Balota, 2008).

Another way in which the difficulty of a sequence of tests can be manipulated is through the difficulty of the questions. In foreign-language vocabulary learning, it is effective to decrease the use of “hints” to the correct word over a sequence of tests (Finley, Benjamin, Hays, Bjork, & Kornell, 2011). An advantage of this technique over the expanding schedule of tests for classroom use is that it does not require complicated scheduling. In both cases, trying to tune the difficulty of ongoing tests to the forgetting that is expected to occur helps to slow the rate of forgetting and enhance long-term memory.

### Tests Improve Inference and Transfer

Thus far, we have only considered how tests benefit a student’s ability to remember material. Of course, remembering what is taught is only a small part of the process of becoming educated in a discipline. Being able to generalize and draw new inferences on the basis of the learned material is critically important if we want students to apply their learning to new situations. And there is evidence that quizzing facilitates the generalization and application of knowledge as well.

In one study (Chan, McDermott, & Roediger, 2006), subjects read a set of passages and were either quizzed on their memory for selected facts from those passages or given extra study time. Prior quizzing enhanced memory for the *non-quizzed* (as well as the quizzed) aspects of the texts when compared with re-studying. The benefit persists after a rather long interval (7 days; Chan, 2010) and even when the final test material is quite distant from the original material being quizzed (Butler, 2010).

In general, these benefits are most pronounced in cases in which the learners took what the authors called a *broad* approach to retrieving responses during the practice quiz: When they thought widely about lots of details relevant to the terms in the question—even if those details were not directly related to the sought-after answer—the benefits of quizzing on related but unquizzed material were most pronounced. So it appears that testing benefits learning in part because of the way that it motivates learners to think about relations among learned facts, and in part because it encourages effective reorganization. Such a result is consistent with the finding reviewed earlier that short-answer tests benefit learners more than do multiple-choice tests, as short-answer tests presumably offer more opportunities for broad thinking. It is also consistent with the well-accepted finding in education that asking and answering “deep questions”—ones that focus on relations, logic, and causation, for example—dramatically benefit student learning (King, 1994).

It is for these reasons that adjunct questions—the thought questions that appear in textbooks alongside the main text—have an overall beneficial effect on learning, even for matters not directly related to those questions (Hamaker, 1986). Furthermore, questions that encourage higher-order thinking

(e.g., reasoning to a new situation) over simple fact retrieval enhance the benefits of adjunct questions yet further. Presumably, one major limitation of adjunct questions—particularly, difficult higher-order ones—is students’ willingness to engage them in the course of reading. Quizzing opportunities in the classroom have the potential to circumvent this inclination.

Testing also boosts our ability to learn concepts. In one example (Jacoby, Wahlheim, & Coane, 2010), subjects were required to learn about different families of birds. One group was given four study blocks with pictures of birds and their family names; the other was given only one study block and three blocks on which they were shown only the picture and asked to retrieve the family name. Feedback was provided after their response. The novel aspect of their procedure was a later test that included entirely new pictures of birds from the same families. The group that had had quizzes was actually better able to sort those new pictures of birds into the appropriate families, indicating that quizzing had done more than simply enhance their memory for the previously studied birds. Rather, it seems to have actually improved their knowledge of the categories that differentiated among the birds. Tests encourage the kind of thinking that is essential not just for retention but also for mentally organizing the acquisition of new material.

One consequence of the recent rapid growth in testing research is that some conclusions are still in flux, and the boundary conditions of some effects of testing are still undiscovered or under debate. The benefits of testing on generalization is an area for which this is particularly true—there are reports of failures to generalize as well (Tran, Rohrer, & Pashler, 2015). What can be stated right now with clarity is that there are likely conditions under which the type of learning engendered by testing will generalize effectively, though the range of those conditions is still under active exploration, and the extent of the benefit is yet unknown. At the very least, remembering is a precondition for generalization—we can be quite sure that conditions under which students remember less of what they have learned are not apt to lead to effective generalization and inference to new situations.

### **Tests Decrease Confusions and Reduce Interference**

So far we have seen that the carefully tailored use of tests can enhance memory for and generalization from previously learned materials. Amazingly, the benefits of tests extend even to materials that are only learned *after* the test! In this section, we review evidence that retrieving information from memory—that is, exactly what a test forces you to do—allows learners to more effectively *segregate* their learning and prevent confusions among topics.

Teachers sometimes use in-class tests to break up a lesson plan. It turns out that this is a good strategy for several reasons,

some of which we have reviewed already. An unexpected benefit is that material that is learned after a test is better remembered. In an experiment by Szpunar, McDermott, and Roediger (2008), subjects learned lists of words and were either quizzed between each list on the preceding list or they completed simple math problems. The important result was revealed on a test for the fifth and final list that they studied: Subjects who had experienced interleaved testing of previously learned lists remembered the final list better. Even though the experience of the subjects was exactly the same from the point of the fifth list onward, the group that had experienced tests on their prior lists remembered more on that final list.

One interpretation is that prior tests may have prevented the materials from the earlier lists from interfering with memory for the final list, but one alternative should be considered as well. Tests may frequently motivate people to study harder in anticipation of those tests.

This interpretation is probably not the whole story, however. For example, it turns out that you can replace those tests with other simple retrieval tasks (such as listing presidents, or states, or types of furniture) that would probably not provide a very strong clue that the final list was to be tested, and the results are the same (Divis & Benjamin, 2014; Pastötter, Schicker, Niedernhuber, & Bäuml, 2011).

The same result has been shown with meaningful passages about animals and short-answer tests—taking short quizzes on the presidents and other topics between passages led to enhanced learning of the material in later passages (Divis & Benjamin, 2014). The mental segregation that comes from a quiz and produces benefits for future learning is not without costs, however. An interleaved quiz also makes the events prior to the quiz more difficult to access on the final test (Divis & Benjamin, 2014). This result is probably due to the fact that segregating two study events also effectively segregates the earlier learning episodes from the time of the test. However, this effect appears to be short lived, and hence probably not of great concern unless the final test occurs very shortly after the quiz.

*Making errors during a test enhances memory for correct answers.* One concern that people have with testing is that test takers will make errors and that the process that leads to those errors will become engrained and will prevent the learner from acquiring the correct solution. Interestingly, this does not appear to be the case; in fact, making errors may even have tangible benefits for learners.

In one representative experiment, Kornell, Hays, and Bjork (2009) asked learners to answer unanswerable questions about made-up events. After the subject provided an answer, learners were given the “correct” answer by the experimenter. Those subjects remembered the “correct” answers better than a group that was given the answer but not given the opportunity to make a mistake prior to being given that answer. Even when tests require people to construct

explanations for scientific observations, being required to produce what were generally erroneous speculations did not reduce subsequent learning from feedback (Kang et al., 2011). A similarly beneficial effect of making errors can also be seen following a *pretesting* phase, where learners answer questions about the material before they study it at all (Richland, Kornell, & Kao, 2009).

All of this is not to say that making errors willy-nilly is good for learning overall. Benefits of making errors are only apparent when the errors are meaningfully related to the learning material. So, having someone estimate the age of a person in a picture, for example, does lead to enhanced memory for the correct answer when that correct answer is given as feedback (McGillivray & Castel, 2010), but guessing a random, unrelated associate to a provided word does not (Huelser & Metcalfe, 2012).

Errors that are made with high confidence are even more likely to be successfully remedied by feedback with the correct answer (Butterfield & Metcalfe, 2001). This result is unexpected because it stands to reason that the errors we make with high confidence are the ones we believe most strongly, and such strong beliefs should be more resistant to change. But this reasoning appears to be wrong. In general, feedback that is surprising draws our attention (Butterfield & Metcalfe, 2006) and improves our memory (Fazio & Marsh, 2009). This effect has been shown in young children (Metcalfe & Finn, 2012), young adults, and older adults (Eich, Stern, & Metcalfe, 2013), indicating generalizability to a wide variety of populations.

### *Our Performance on Tests Tunes Our Knowledge About What We Do and Do Not Know*

One of the reasons that tests are unappealing to some students and to their overweening parents is that tests fairly reveal what we do and do not know. This feedback can violate the positive feelings we hold about ourselves and our abilities, which are often inappropriately optimistic, especially in the classroom (Hacker, Bol, Horgan, & Rakow, 2000). This violation causes students to rate instructors more poorly (Isley & Singh, 2005) and to generate complicated but unsupported theories about supposed learning styles that their classrooms are failing to support (Pashler, McDaniel, Rohrer, & Bjork, 2009). What is of particular concern is the way that such inappropriately tuned self-assessments influence study behavior.

There are two ways in which having a poor calibration between what we know and what we *think* we know can be harmful. The first concern is that inappropriate overconfidence will lead us to study less than is warranted (Bandura, 1993). Why would a student who feels that they have mastered the material continue to study? The second is that poor insight into what we have or have not yet mastered can lead to poor decisions about how to allocate our study when we

do choose to study. Spending additional time on already mastered aspects of the curriculum or spending very little time on poorly mastered aspects is a non-optimal use of learners' time.

**Confidence.** Overconfidence in one's abilities and judgments is ubiquitous across domains (e.g., Fischhoff, Slovic, & Lichtenstein, 1977; Klayman, Soll, Gonzalez-Vallejo, & Barlas, 1999). Experts in a domain are not immune to this bias; on the contrary, they may be even more susceptible. Soccer experts, for example, have been shown to predict the outcome of World Cup matches with the same accuracy as non-experts, but with much higher confidence (Andersson, Edman, & Ekman, 2005). It is easy to see what the costs of such overconfidence might be: a failure to gather more evidence prior to committing to a decision, an unwillingness to consider the opinions of others, making inappropriate wagers, and the like.

Analogous costs are apparent for the overconfident student: insufficient study time, poor prioritization of study techniques, and so on. And, indeed, students who exhibit more overconfidence in their assessment of mastery on a set of definitions do indeed reveal poorer performance in later exams on those materials (Dunlosky & Rawson, 2012). On the other side of the coin, students with higher grade-point averages (GPAs) exhibit lower overconfidence when predicting performance in an upcoming exam (Grimes, 2002).

There are techniques that are known to reduce overconfidence. Most relevant for our discussions here is the role of feedback in ongoing learning and testing. Expert forecasters who are forced to directly compare their predictions with the outcomes of the events they are trying to predict often exhibit exquisitely tuned calibration. A terrific example is provided by weather forecasters, who get all kinds of feedback from bosses and angry citizens when their predictions are wrong. Consequently, weather forecasting is now an exceptionally accurate endeavor (Murphy & Winkler, 1984). Contrast this with sports or political forecasting, where the outcomes are available but often discounted when they fail to confirm one's predictions or ignored entirely in the hubbub of the next major event (Silver, 2012).

Tests provide the opportunity for students to tune their confidence in their understanding and mastery of course materials to appropriate levels. Students who receive immediate feedback on the accuracy of their responses by a computerized testing system reveal much more enhanced calibration of confidence than students who do not receive feedback (Zakay, 1992). Although standardized tests do not often force the students to directly compare their confidence in material prior to the exam with the outcome of that exam, later review of that exam can serve that function, as can exams that require students to make decisions about which answers they choose to submit for grading (e.g., Higham, 2007) or how precise an answer to submit (Higham, 2013).

Students who experience multiple cycles of studying material, making judgments about their ongoing learning, and taking a test exhibit improved judgment accuracy over cycles (Kelemen, Winningham, & Weaver, 2007). In addition, computerized learning environments that include the opportunity to take tests and to assess one's performance reveal an advantage on later tests over control conditions that do not include such forced assessments (Metcalf, Kornell, & Son, 2007).

*Tests inform us of what topics are important and where our learning is deficient.* One of the most direct ways in which tests promote learning is by motivating students to study. The benefits of this effect can be controversial when it is believed that the test measures unimportant skills or when teachers focus on the test to the exclusion of other materials, two common criticisms of the current standardized tests for the Common Core. But the curriculum for the Common Core, as well as its attendant tests, is fluid and likely to experience considerable development. Students who take regular quizzes in the classroom are more likely to attend unrequired meetings (Fitch, Drucker, & Norton, 1951) and exhibit better class attendance (Wilder, Flood, & Stromsnes, 2001), both of which are known to increase student achievement. Moreover, tests with a clear agenda can focus teachers' and students' activities onto materials that are broadly considered to be valuable.

Students learn very rapidly from tests how to direct their study to important materials. They learn to ignore materials that are not likely to be tested, to relate materials in such a way that conforms to the expected nature of the test (Finley & Benjamin, 2012), and to study more for tests that they expect to be difficult (Meyer, 1936).

Tests also sharpen focus on important but unmastered materials (Thiede & Dunlosky, 1999). Having the opportunity to attempt to retrieve information from memory on a test—where the learner does not have easy access to the answer—is highly diagnostic of one's level of mastery for that material (Benjamin, Bjork, & Schwartz, 1998) and consequently highly related to the success of future attempts to retrieve the material (Nelson & Dunlosky, 1991). Study following a test is more effective in part because it enhances the extent to which learners spend more time with the more difficult, unmastered materials (Soderstrom & Bjork, 2014), a strategy that is highly effective (Tullis & Benjamin, 2011).

*Tests help us learn which learning techniques are effective and which are ineffective.* One bottleneck to effective student learning is the widespread use of poor learning techniques. Many of the techniques that have been identified as highly effective in basic research on learning and memory (for reviews, see, for example, Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Pashler et al., 2007) are ones that are dispreferred by students. Most people, for example, eschew distributing practice (Baddeley & Longman, 1978),

interleaving multiple to-be-learned skills (Simon & Bjork, 2001), and—most poignantly for this review—testing (Karpicke, 2009)!

When we take a test, we discover which of our study endeavors have been successful and which have not been. This appears to be particularly true when learners make judgments about those techniques (or about materials learned with those techniques) prior to the test (e.g., Begg, Duft, Lalonde, Melnick, & Sanvito, 1989; Benjamin, 2003; Fiechter, Benjamin, & Unsworth, in press), and then have the opportunity to see the consequences of those varying techniques play out in memory performance. The benefits of quizzing oneself are, for example, apparent on a later test, and learners change their assessments of that technique when guided to observe the gap between their assessments of how testing will influence memory and their actual later performance (Tullis, Finley, & Benjamin, 2013).

Overall, it has been found in educational settings that opportunities for self-assessment enhance performance in the classroom (Dochy, Segers, & Sluijmsmans, 1999). In combination with performance tests, self-assessments can help students gain a better appreciation for (a) their overall level of competence, (b) what they do and do not currently know, and (c) which study techniques are serving them most effectively. These techniques comprise a skill that has been recognized as critical for effective learning and transfer for more than half a century: “learning to learn” (Postman, 1969).

## The Effective Use of Tests Within Standardized Testing

We have summarized here only a small portion of the voluminous literature within psychology, education, and forecasting, showing that testing can promote durable and generalizable learning in a wide variety of important ways. Standardized tests are critical for the assessment of students' progress, and of our nation's progress toward a standard that will ensure international competitiveness for our graduates. It is certain to be a central part of educational reform in the United States, but that does not mean that we cannot conceive of a broader role for testing in education.

Well-designed tests can be part of the solution to improved educational outcomes. The literature reviewed here suggests some ways in which traditional standardized tests can be modified to take greater advantages of these qualities. SBAC and PARCC are both using computer-adaptive systems for testing, a quality that enables certain interventions that will boost the ways in which tests serve as learning events. Knowing that tests serve a role in learning, and not merely assessment, might allay some of the major concerns that students (Strauss, 2015), school administrators (Perez & Rado, 2015), and governments (Harris, 2015) have with standardized tests as they are currently implemented. For tests to promote learning as greatly as they could, however, the character

of those tests, and the procedures used in their administration, may need to change. We end our discussion with some speculative thoughts on how “high-stakes” testing itself could be reformed in line with the suggestions from the experimental evidence described above.

1. Tests should encourage broad and deep thinking, as well as explicit direct retrieval of facts, to facilitate the generalization and transfer of knowledge. Education requires a mix of explicit direct learning and skill learning, much of which we want to generalize to new domains. Many tests emphasize material that can be memorized, in part because it can be more easily evaluated. When thinking about the trade-off between the breadth of a test question and how easy it will be to score the answer, consideration should be given to the beneficial effects such broad questions are known to have.
2. Having a larger number of tests, rather than a single test at the end of a school year, can have multiple benefits. From a measurement perspective, multiple tests provide multiple snapshots of a student or of a classroom, and as such, are more likely to reflect student achievement and less likely to be influenced by a single bad day. From a psychological perspective, multiple tests enhance learning and render each test a lower-stakes event, which may decrease student and teacher anxiety. Moreover, administering tests well before the end of the school year can potentially allow the results to guide changes in students’ and teachers’ learning strategies and efforts during the same year.
3. Tests can be designed to encourage self-assessment. Tests that require students to make decisions about their confidence in provided answers may already serve this function. When a test promotes a more accurate view of what a student does and does not know, the effectiveness of future study activities can be improved.
4. The more widespread use of adaptive testing procedures can ensure that students taking a test are appropriately challenged throughout. Students who confront abject failure in a test are unlikely to experience many of the benefits reviewed here, and are more likely to suffer distress and perform poorly in exams (Fincham, Hokoda, & Sanders, 1989). On the other hand, students who find an exam too easy will not experience as much benefit to their enduring memory of course material compared with students who are challenged somewhat. A well-designed adaptive testing regimen can ensure that each student is challenged appropriately.
5. The incorporation of feedback into tests is central to many of the cognitive benefits tests can provide.

With computerized testing, immediate feedback is possible, though consideration must be given as to how to do so without compromising the assessment purposes of the test. Later review of exam materials, perhaps even in the classroom, can be an effective way of ensuring that some of the benefits of tests can be enjoyed by the students who take them.

When considering the multitude of changes that the U.S. educational system is currently undergoing, it is critical that we keep multiple targets in our sight. The traditional view is that there is a separation between the acquisition of knowledge and skills and the evaluation of a student’s mastery of the curriculum. Both are important goals, and current research indicates that tests can facilitate progress on both fronts. Keeping in mind the ways that tests can be fruitfully used to enhance education, rather than simply measure it, allows us to take a broader view of the role of standardized tests in modern education policy.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### References

- Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, 2, 396-408.
- Andersson, P., Edman, J., & Ekman, M. (2005). Predicting the World Cup 2002 in soccer: Performance and confidence of experts and non-experts. *International Journal of Forecasting*, 21, 565-576.
- Baddeley, A. D., & Longman, D. J. A. (1978). The influence of length and frequency of training session on the rate of learning to type. *Ergonomics*, 21, 627-635.
- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, 28, 117-148.
- Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C. L. C. (1991). Effects of frequent classroom testing. *Journal of Educational Research*, 85, 89-99.
- Benjamin, A. S. (2003). Predicting and postdicting the effects of word frequency on memory. *Memory & Cognition*, 31, 297-305.
- Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory & Language*, 28, 610-632.
- Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metacognitive index. *Journal of Experimental Psychology: General*, 127, 55-68.
- Berlyne, D. E. (1966). Conditions of prequestioning and retention of meaningful material. *Journal of Educational Psychology*, 57, 128-132.



- Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1118-1133.
- Butterfield, B., & Metcalfe, J. (2001). Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1491-1494.
- Butterfield, B., & Metcalfe, J. (2006). The correction of errors committed with high confidence. *Metacognition and Learning*, *1*, 1556-1623.
- Camp, C. J. (2006). Spaced retrieval: A model for dissemination of a cognitive intervention for persons with dementia. In D. K. Attix & K. A. Welsh-Bohmer (Eds.), *Geriatric neuropsychology: Assessment and intervention* (pp. 275-292). New York, NY: Guilford Press.
- Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of US history facts. *Applied Cognitive Psychology*, *23*, 760-771.
- Carpenter, S., Pashler, H., Wixted, J., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition*, *36*, 438-448.
- Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, *20*, 632-642.
- Cattell, J. M. (1890). Mental tests and measurement. *Mind*, *15*, 373-380.
- Chan, J. C. K. (2010). Long-term effects of testing on the recall of nontested materials. *Memory*, *18*, 49-57.
- Chan, J. C. K., McDermott, K. B., & Roediger, H. L. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*, *135*, 553-571.
- Clement, S., & Brown, E. (2015, February 20). Poll: Widespread misperceptions about the Common Core standards. *Washington Post*. Retrieved from [http://www.washingtonpost.com/local/education/poll-widespread-misperceptions-about-the-common-core-standards/2015/02/20/9e844310-b926-11e4-aa05-1ce812b3fdd2\\_story.html](http://www.washingtonpost.com/local/education/poll-widespread-misperceptions-about-the-common-core-standards/2015/02/20/9e844310-b926-11e4-aa05-1ce812b3fdd2_story.html)
- Divis, K. M., & Benjamin, A. S. (2014). Retrieval speeds context fluctuation: Why semantic generation enhances later learning but hinders prior learning. *Memory & Cognition*, *42*, 1049-1062.
- Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, *24*, 331-350.
- Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction*, *22*, 271-280.
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*, 4-58.
- Eich, T. S., Stern, Y., & Metcalfe, J. (2013). The hypercorrection effect in younger and older adults. *Aging, Neuropsychology, and Cognition*, *20*, 511-521.
- Fazio, L. K., & Marsh, E. J. (2009). Surprising feedback improves later memory. *Psychonomic Bulletin & Review*, *16*, 88-92.
- Fiechter, J. L., Benjamin, A. S., & Unsworth, N. (in press). The metacognitive foundations of effective remembering. In J. Dunlosky & S. K. Tauber (Eds.), *Oxford handbook of metamemory*. New York, NY: Oxford University Press.
- Fincham, F. D., Hokoda, A., & Sanders, R. (1989). Learned helplessness, test anxiety, and academic achievement: A longitudinal analysis. *Child Development*, *60*, 138-145.
- Finley, J. R., & Benjamin, A. S. (2012). Adaptive and qualitative changes in encoding strategy with experience: Evidence from the test-expectancy paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 632-652.
- Finley, J. R., Benjamin, A. S., Hays, M. J., Bjork, R. A., & Kornell, N. (2011). Benefits of accumulating versus diminishing cues in recall. *Journal of Memory and Language*, *64*, 289-298.
- Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 552-564.
- Fitch, M. L., Drucker, A. J., & Norton, J. A. (1951). Frequent testing as a motivating factor in large lecture classes. *Journal of Educational Psychology*, *42*, 1-20.
- Forlano, G. (1936). *School learning with various methods of practice and rewards*. New York, NY: Teachers College, Columbia University.
- Fritz, C. O., Morris, P. E., Nolan, D., & Singleton, J. (2007). Expanding retrieval practice: An effective aid to preschool children's learning. *The Quarterly Journal of Experimental Psychology*, *60*, 991-1004.
- Galton, F. (1879). Psychometric experiments. *Brain*, *2*, 149-162.
- Gates, A. I. (1917). Recitation as a factor in memorizing. *Archives of Psychology*, *6*(40).
- Glover, J. A. (1989). The "testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, *81*, 392-399.
- Grimes, P. W. (2002). The overconfidence principles of economics student: An examination of a metacognitive skill. *The Journal of Economic Education*, *33*, 15-30.
- Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology*, *92*, 160-170.
- Hamaker, C. (1986). The effects of adjunct questions on prose learning. *Review of Educational Research*, *56*, 212-242.
- Harris, E. A. (2015, March 1). As common core testing is ushered in, parents and students opt out. *The New York Times*. Retrieved from [http://www.nytimes.com/2015/03/02/nyregion/as-common-core-testing-is-ushered-in-parents-and-students-opt-out.html?hp&action=click&pgtype=Homepage&module=second-column-region&region=top-news&WT.nav=top-news&\\_r=1](http://www.nytimes.com/2015/03/02/nyregion/as-common-core-testing-is-ushered-in-parents-and-students-opt-out.html?hp&action=click&pgtype=Homepage&module=second-column-region&region=top-news&WT.nav=top-news&_r=1)
- Hellyer, S. (1962). Frequency of stimulus presentation and short-term decrement in recall. *Journal of Experimental Psychology*, *64*, 650.
- Higham, P. A. (2007). No special K! A signal detection framework for the strategic regulation of memory accuracy. *Journal of Experimental Psychology: General*, *136*, 1-22.
- Higham, P. A. (2013). Regulating accuracy on university tests with the plurality option. *Learning and Instruction*, *24*, 26-36.
- Huelser, B. J., & Metcalfe, J. (2012). Making related errors facilitates learning, but learners do not know it. *Memory & Cognition*, *40*, 514-527.
- Isley, P., & Singh, H. (2005). Do higher grades lead to favorable student evaluations? *Journal of Economic Education*, *36*, 29-42.

- Jacoby, L. L., Wahlheim, C. N., & Coane, J. H. (2010). Test-enhanced learning of natural concepts: Effects on recognition memory, classification, and metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 1441-1451.
- Kang, S. H. K., Lindsey, R., Mozer, M. C., & Pashler, H. (2014). Retrieval practice over the long term: Should spacing be expanding or equal-interval? *Psychonomic Bulletin & Review*, 21, 1544-1550.
- Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, 19, 528-558.
- Kang, S. H. K., & Pashler, H. (2014). Is the benefit of retrieval practice modulated by motivation? *Journal of Applied Research in Memory and Cognition*, 3, 183-188.
- Kang, S. H. K., Pashler, H., Cepeda, N. J., Rohrer, D., Carpenter, S. K., & Mozer, M. C. (2011). Does incorrect guessing impair fact learning? *Journal of Educational Psychology*, 103, 48-59.
- Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General*, 138, 469-486.
- Kelemen, W. L., Winingham, R. G., & Weaver, C. A. (2007). Repeated testing sessions and scholastic aptitude in college students' metacognitive accuracy. *European Journal of Cognitive Psychology*, 19, 689-717.
- King, A. (1994). Guiding knowledge construction in the classroom: Effects of teaching children how to question and how to explain. *American Educational Research Journal*, 31, 338-368.
- Klayman, J., Soll, J. B., Gonzalez-Vallejo, C., & Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational Behavior and Human Decision Processes*, 79, 216-247.
- Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 989-998.
- Landauer, T. K., & Bjork, R. A. (1978). Optimal rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625-632). London, England: Academic Press.
- LaPorte, R. E., & Voss, J. F. (1975). Retention of prose materials as a function of postacquisition testing. *Journal of Educational Psychology*, 67, 259-266.
- Logan, J. M., & Balota, D. A. (2008). Expanded versus equal interval spaced retrieval practice: Exploring different schedules of spacing and retention interval in younger and older adults. *Aging, Neuropsychology, and Cognition*, 15, 257-280.
- McDaniel, M. A., Agarwal, P. K., Huelsner, B. J., McDermott, K. B., & Roediger, H. L. (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology*, 103, 399-414.
- McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, 19, 494-513.
- McDaniel, M. A., & Fisher, R. P. (1991). Tests and test feedback as learning sources. *Contemporary Educational Psychology*, 16, 192-201.
- McGillivray, S., & Castel, A. D. (2010). Memory for age-face associations: The role of generation and schematic support. *Psychology and Aging*, 25, 822-832.
- Metcalfe, J., & Finn, B. (2012). Hypercorrection of high confidence errors in children. *Learning and Instruction*, 22, 253-261.
- Metcalfe, J., Kornell, N., & Son, L. K. (2007). A cognitive-science based programme to enhance study efficacy in a high and low risk setting. *European Journal of Cognitive Psychology*, 19, 743-768.
- Meyer, G. (1936). The effects of recall and recognition on the examination set in the classroom. *Journal of Educational Psychology*, 27, 81-99.
- Mulligan, N. W. (2005). Total retrieval time and hypermnesia: Investigating the benefits of multiple recall tests. *Psychological Research*, 69, 272-284.
- Murphy, A. H., & Winkler, R. L. (1984). Probability forecasting in meteorology. *Journal of the American Statistical Association*, 79, 489-500.
- Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect." *Psychological Science*, 2, 267-270.
- Nelson, T. O., & Vining, S. K. (1978). Effect of semantic versus structural processing on long-term retention. *Journal of Experimental Psychology: Human Learning and Memory*, 4, 198-208.
- Pan, S. C., Pashler, H., Potter, Z. E., & Rickard, T. C. (2015). Testing enhances learning across a range of episodic memory abilities. *Journal of Memory and Language*, 83, 53-61.
- Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). *Organizing instruction and study to improve student learning (NCER 2007-2004)*. Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.
- Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. A. (2009). Learning styles: Concepts and evidence. *Psychological Science in the Public Interest*, 9, 105-119.
- Pastötter, B., Schicker, S., Niedernhuber, J., & Bäuml, K.-H. T. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 287-297.
- Pearson. (2014). *The learning curve*. Retrieved from <http://thelearningcurve.pearson.com/reports/the-learning-curve-report-2014>
- Perez, J., & Rado, D. (2015, March 2). CPS reverses course, will administer controversial test. *Chicago Tribune*. Retrieved from <http://www.chicagotribune.com/news/chi-cps-common-core-test-20150302-story.html#page=1>
- Postman, L. (1969). Experimental analysis of learning to learn. In G. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 3; pp. 241-297). New York, NY: Academic Press.
- Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied*, 15, 243-257.
- Roediger, H. L., 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. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249-255.
- Schacter, D. L., Rich, S. A., & Stampp, M. S. (1985). Remediation of memory disorders: Experimental evaluation of the spaced retrieval technique. *Journal of Clinical and Experimental Neuropsychology, 7*, 79-96.
- Silver, N. (2012). *The signal and the noise: Why so many predictions fail—But some don't*. New York, NY: Penguin Press.
- Simon, D. A., & Bjork, R. A. (2001). Metacognition in motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 907-912.
- Soderstrom, N. C., & Bjork, R. A. (2014). Testing facilitates the regulation of subsequent study time. *Journal of Memory and Language, 73*, 99-115.
- Storm, B. C., Bjork, R. A., & Storm, J. C. (2010). Optimizing retrieval as a learning event: When and why expanding retrieval practice enhances long-term retention. *Memory & Cognition, 38*, 244-253.
- Strauss, V. (2015, March 7). A 16-year-old takes the new PARCC exam. Here's her disturbing report. *Washington Post*. Retrieved from <http://www.washingtonpost.com/blogs/answer-sheet/wp/2015/03/07/a-16-year-old-takes-the-new-parcc-exam-heres-her-disturbing-report/>
- Szpunar, K. K., McDermott, K. B., & Roediger, H. L. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 1392-1399.
- Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1024-1037.
- Tran, R., Rohrer, D., & Pashler, H. (2015). Retrieval practice: The lack of transfer to deductive inferences. *Psychonomic Bulletin & Review, 22*, 135-140.
- Tullis, J. G., & Benjamin, A. S. (2011). On the effectiveness of self-paced learning. *Journal of Memory and Language, 64*, 109-118.
- Tullis, J. G., Finley, J. R., & Benjamin, A. S. (2013). Metacognition of the testing effect: Guiding learners to predict the benefits of retrieval. *Memory & Cognition, 41*, 429-442.
- Wheeler, M. A., Ewers, M., & Buonanno, J. F. (2003). Different rates of forgetting following study and test trials. *Memory, 11*, 571-580.
- Wilder, D. A., Flood, W. A., & Stromsnes, W. (2001). The use of random extra credit quizzes to increase student attendance. *Journal of Instructional Psychology, 28*, 117-120.
- Wilson, B. A., Baddeley, A., Evans, J., & Sheil, A. (1994). Errorless learning in the rehabilitation of memory impaired people. *Neuropsychological Rehabilitation, 4*, 307-326.
- World Economic Outlook Database. (2014, October 7). International Monetary Fund. Retrieved from: <https://www.imf.org/external/pubs/ft/weo/2013/02/weodata/index.aspx>
- Wurman, Z. (2015, June 7). He's dead, Jim: Why common core is a goner and just doesn't know it yet. *Breitbart*. Retrieved from <http://www.breitbart.com/big-government/2015/06/07/hes-dead-jim-why-common-core-is-a-goner-and-just-doesnt-know-it-yet/>
- Zakay, D. (1992). The influence of computerized feedback on overconfidence in knowledge. *Behavior & Information Technology, 11*, 329-333.
- Zaromb, F. M., & Roediger, H. L. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition, 38*, 995-1008.