

Reflections on the Resurgence of Interest in the Testing Effect

Henry L. Roediger, III¹, and Jeffrey D. Karpicke²

¹Department of Psychological and Brain Sciences, Washington University in St. Louis, and ²Department of Psychological Sciences, Purdue University

Abstract

We discuss the findings from our 2006 article in *Psychological Science* on the testing effect and describe how the project arose. The testing effect (or retrieval-practice effect) was first reported in the experimental literature about a century before our article was published, and the effect had been replicated (and sometimes discovered anew) many times over the years. Our experiments used prose materials (unlike most prior research) and produced a more powerful effect than prior research even though we used a conservative control condition for comparison. In our discussion, we drew out possible implications for educational practice. We also reported that students in the experiment could not predict the effect; this lack of metacognitive awareness represented a new finding in this context. In a companion article the same year, we provided an historical review of the testing effect. We believe the synergistic effect of the two articles accounts in part for the resurgence in interest in this phenomenon and its application in educational settings.

Keywords

cognition, memory, application, education

If someone had suggested in 2006, when we published our article (Roediger & Karpicke, 2006b) in *Psychological Science*, that a decade later we would be asked to write about why it had such great impact, we would have dismissed the idea. We certainly thought it was a worthy and interesting contribution, but one of the 30 most-cited articles in APS journals during the first 30 years of the organization? Highly unlikely. After all, as one reviewer of our initial submission pointed out, the pattern we reported in our article had already appeared at least twice in the literature (albeit in weaker form), and so ours was in some sense a conceptual replication. This reviewer suggested it might not be suitable for so august a journal as *Psychological Science* because of this fact. Indeed, we had debated which journal was most appropriate before we submitted it, but we finally decided on *Psychological Science* because we obtained much larger effects than had been reported previously, we used more educationally relevant material, and we reported a new finding regarding metacognition.

But we are getting ahead of ourselves. The Editor has asked us to address five questions in crafting our article, and we follow his advice.

What Main Scientific Psychological Question Did the Article Address?

We asked whether students would learn more from a prose passage after reading it once if they took a test on the passage rather than reading the passage again (the control condition). In cognitive psychology, as in education, studying is usually credited with how we learn, and testing is thought merely to assess what learning has occurred via study. Yet experiments dating back to 1909, mostly with word lists, have shown that recitation (via a test) can also be a potent factor in learning (Abbott, 1909). As noted, we used prose materials in our experiments and compared taking a test after a first reading as a learning activity with rereading the passage as the control condition, because rereading is how students usually study (Karpicke, Butler, &

Corresponding Author:

Henry L. Roediger, III, Department of Psychological and Brain Sciences, Washington University in St. Louis, One Brookings Dr., St. Louis, MO 63130-4899
E-mail: roediger@wustl.edu

Roediger, 2009). Of course, when restudying, students are reexposed to 100% of the information, whereas when taking a test, they are reexposed only to the percentage of the material they can recall (about 70% in our experiments). Thus, the deck is stacked against finding a benefit from retrieval practice via testing when rereading of the material is used as the control condition. We also used repeated tests compared with repeated study episodes of the passages; again the tests occurred without feedback. If we obtained a positive effect of testing, would more tests enhance it?

What Was the Article About? Provide a Summary

We performed two experiments, providing a replication of our basic finding. We describe Experiment 2 here, which consisted of a 3 (learning conditions) \times 2 (retention interval) design. The learning conditions were denoted SSSS, SSST, and STTT (S = study; T = test). In all three conditions, student subjects were given a passage to learn that was about 250 words in length. In the SSSS condition, students read the passage repeatedly in four separate 5-min blocks; because the passage was short, subjects in this condition read it about 14 times in total. In the SSST condition, students read the passage during three 5-min blocks (reading it about 10 times) and then took a single 5-min test that asked them to recall as much of the passage as possible; they recalled 70% of the ideas in the passage (the SSST conditions). In the STTT condition, subjects read the passage in only one 5-min period (about 3.5 times, on average) and then recalled it three times (recovering about 70% of the ideas each time). After the learning phase, subjects rated the passage on several dimensions, including how well they thought they would remember it a week later. Then half the subjects were given a final test after only 5 min whereas the other half were dismissed and came back to the lab a week later for their final test.

The results are shown in Figure 1 (Fig. 2 in the original article) and are easy to describe: On the immediate test, the greater the amount of studying, the better the recall. However, a week later the exact opposite pattern held: The more retrieval practice (and consequently the less studying) subjects did during the learning session, the better they recalled the information. This retrieval practice effect and the cross-over interaction had been reported previously, in experiments with word lists, although our findings with prose provided much larger differences (we discuss the reasons later). One new finding concerned students' metacognitions: When students rated how well they thought they would recall the passage after a week, those in the SSSS condition provided the highest ratings (on a 7-point scale) and

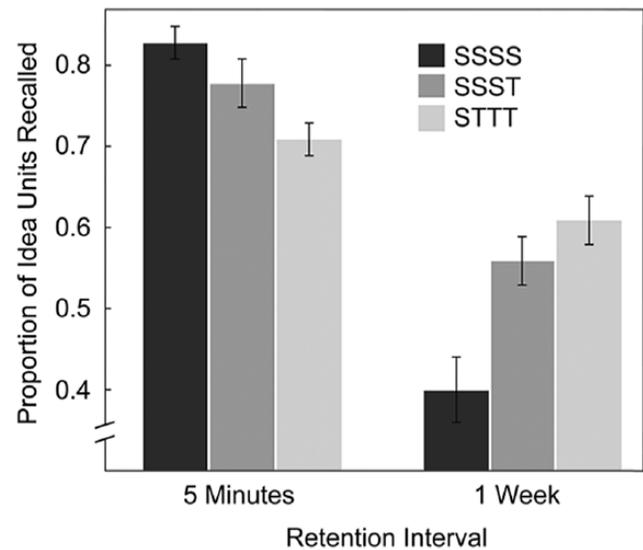


Fig. 1. Mean proportion of idea units recalled on the final test after a 5-min or 1-week retention interval as a function of learning condition (SSSS, SSST, or STTT) in Experiment 2. The labels for the learning conditions indicate the order of study (S) and test (T) periods. Error bars represent standard errors of the means.

those in the STTT condition predicted they would recall the worst, exactly the opposite of the actual outcome after a week.

From another perspective, the experiment shows that cramming (repeated reading) can lead to good performance on an immediate test but that students will forget much of what they once knew after a week. Because students generally report that they study by rereading (Karpicke et al., 2009), the rapid forgetting after repeated studying may explain why so many professors complain that students retain little of what they learn. The Roediger and Karpicke (2006b) experiments show the power of retrieval practice, as did other experiments we reported shortly thereafter (Karpicke & Roediger, 2007, 2008). Numerous other researchers also reported impressive effects of testing or retrieval practice (e.g., Soderstrom, Kerr, & Bjork, 2016), and in fact such studies have been reported sporadically during the 20th century. Rowland (2014) provided a meta-analytic review of the testing-effect literature (see somewhat different meta-analyses by Adesope, Trevisan, & Sundararajan, 2017; Schwieren, Barenberg, & Dutke, 2017). The effect is highly reliable.

What Do You See as the Main Contribution of the Article to Psychological Science and Society?

First, we showed the power of retrieval practice (or testing) with materials that were at least somewhat relevant to education. The research was soon extended

into more realistic classroom simulations (e.g., Butler & Roediger, 2007) and then into actual educational situations. The promise of retrieval practice as an effective learning technique has been borne out repeatedly in a variety of settings: university classes (e.g., Butler, Marsh, Slavinsky, & Baraniuk, 2014; Lyle & Crawford, 2011; McDaniel, Anderson, Derbish, & Morrisette, 2007), medical education (e.g., Larsen, Butler, & Roediger, 2009), middle school classrooms (e.g., McDermott, Agarwal, D'Antonio, Roediger, & McDaniel, 2014), and elementary school classrooms (Karpicke, Blunt, & Smith, 2016). Retrieval practice as a technique has been adopted by many others in various venues and nearly always seems to boost performance. Karpicke and Blunt (2011) showed that retrieval practice enhanced learning to a greater extent than the use of concept maps, and in a later experiment, they showed that combining the techniques—retrieval in the form of a concept map—also produces improved learning (Blunt & Karpicke, 2014).

Regarding another applied implication, the data shown in Figure 1 conform to the pattern that Bjork (1994) described as exemplifying desirable difficulties in learning. Research in several arenas indicates that variables that slow initial learning and make it feel more difficult may provide a beneficial effect on long-term retention. Having tests interspersed with episodes of study slows initial learning (as seen on the 5-min test in Fig. 1) but enhances long-term retention (as seen in 1-week test). Other variables, such as spacing and interleaving of practice, have the same effect as retrieval practice: enhancing long-term learning but impairing performance in the short term (see Kang, 2017; Putnam, Nestojko, & Roediger, 2017).

One reason that learners, as well as teachers and trainers, misunderstand what learning strategies are effective in the long term is that they naturally focus on short-term performance, the techniques that bring learning up to speed quickly. Knowing what conditions foster good long-term recall is hard, short of measuring it, and teachers, trainers, and students rarely do. Recall too that subjects in our experiments who experienced the repeated study condition (SSSS) predicted that they would recall the passage better a week later than those who read it much less but practiced retrieving it three times (the STTT condition), yet the results showed the opposite. Tests have multiple benefits in enhancing learning in addition to the direct effect of retrieval practice (see Roediger, Putnam, & Smith, 2011).

How Did You Get the Idea for the Article?

In a sense, getting the idea was easy, because other researchers used a design like ours before we did.

Thompson, Wenger, and Bartling (1978) gave subjects a word list to study. In one condition, the subjects studied the list multiple times but took no tests; in the other condition, subjects studied the list one time and then took three successive tests. Subjects in the repeated-study condition recalled more words on the immediate test, but this difference disappeared on the delayed test. The repeated-testing group recalled 3% more words after 48 hr, although of course the difference was not significant (see too Hogan & Kintsch, 1971). In an influential chapter, R. A. Bjork (1975) emphasized the fact that retrieval can serve as a “memory modifier.”

Somewhat later, Wheeler, Ewers, and Buonanno (2003) replicated the findings of Thompson et al. (1978), and in one of their experiments, they sought to show an advantage of the repeated-testing condition relative to the repeated-study condition using a 1-week retention interval (rather than equivalent performance). They did so, and the outcome was statistically significant, but it was numerically rather small (about a 5% advantage for the repeated-testing condition compared with the repeated-study condition).

These experiments (and others) showed that repeated testing could, after a delay, produce recall as good as or slightly better than repeated studying. But could the effect be magnified and thus be of practical use? Low initial levels of recall in the aforementioned studies seemed a likely factor in producing a small testing effect. Testing cannot provide a benefit for material that is not retrieved unless feedback is given on the tests (Kang, McDermott, & Roediger, 2007). Feedback was not given in the word-list experiments just described, and it was not provided in our experiments—but recall of idea units was relatively high at 70% in our studies. Again, to obtain a positive effect of retrieval practice, a necessary condition is to retrieve the material (unless feedback is given).

In 1992, Wheeler and Roediger published an article that was about a different phenomenon, but it included various amounts of retrieval practice in recall of pictures before a delayed test. Performance on the initial tests was reasonably good. One incidental finding was that testing subjects three times produced a larger effect than a single test on recall a week later, and the single test created better retention than no testing at all. In addition, three tests nearly eliminated forgetting a week later! The effect was much greater than that seen in word-list experiments. Although this outcome was only briefly mentioned in the Wheeler and Roediger (1992) article, Roediger found himself thinking back to it repeatedly and wondering whether recall on tests could be boosted by using more meaningful materials. Could repeated testing be shown to be much greater than

repeated studying on a delayed test? The Wheeler and Roediger (1992) experiment did not include a restudy condition.

Jeff Karpicke entered graduate school in the fall of 2002 and joined the Roediger lab. Roediger suggested that they might collaborate in pursuing how testing (or retrieval practice) affects later memory, giving him several articles to read including the ones mentioned in this section. (Karpicke retains a vivid memory of this conversation; Roediger does not.) We then combined the various ideas outlined above using materials that were at least closer to the type used in education, because we hoped to produce results relevant for education. The experiments accomplished the objective of showing how powerful retrieval practice via testing could be relative to repeated studying. Even though weak patterns such as ours had been reported in the literature, the reviewers and editors were won over by our new work (although it was something of a close call). Of course, our article was published in an era when replication was viewed as a cause for rejection rather than celebration, as is the case now. Very few variables create cross-over interactions as a function of retention interval in the study of human memory.

We are focusing on our own research here, but many excellent articles on the testing effect preceded ours (e.g., Carrier & Pashler, 1992, among many others) or came out at about the same time (Carpenter & DeLosh, 2005). And since 2006, the floodgates have opened, and today there are hundreds of articles on retrieval practice, both in the lab and in applied contexts. The effect has been widely replicated, but as with any effect, boundary conditions do exist (see the immediate test data in Fig. 1). Our results have been replicated both directly and conceptually. Einstein, Mullet, and Harrison (2012) converted the experiment into a teaching demonstration to convince students to use retrieval practice. The basic effect is easy to obtain, even in less than ideal circumstances.

Why Has the Article Had Such Impact?

Good question, and we can only speculate. Beginning in the early 2000s, many cognitive psychologists who study learning and memory turned their attention to the educational implications of their work. Despite more than a century of research on learning and memory, the translation of findings from researchers' labs to classroom practice is a rare occurrence (Roediger, 2013). In 2002, the U.S. Education Sciences Reform Act was passed by the U.S. Congress, creating the Institute for Education Science, which supported translational research to improve education. In addition, the No Child Left Behind Act required that education be based

on scientifically grounded research, and the evidence-based practice of education became a rallying cry in some quarters. Perhaps for these reasons (and doubtless others), numerous cognitive researchers turned their attention to issues in education and joined forces with educational psychologists who had worked on these issues for years. The James S. McDonnell Foundation provided a 10-year collaborative-activity grant that brought together 11 researchers spread across 6 universities and supported their research on applying cognitive psychology to enhance educational practice.

Experiments showing the power of retrieval practice in enhancing learning led many researchers to explore both the theoretical reasons for the effect as well its applied implications. Of course, the testing effect was not new, but much of the prior research was conducted with word lists, and our 2006 article used more educationally relevant material. In addition, we showed that students did not predict the effect; rather, they predicted that repeated studying would help them more than repeated testing. And the idea that tests can help learning seemed novel (and implausible) to many people. So, the claim that tests could serve a useful educational function besides assessment and the assignment of grades attracted attention. One journalist told us that the basic finding (testing or retrieval leading to better long-term recall than studying) was "deeply counterintuitive."

Edwin G. Boring, the eminent historian of psychology, often wrote of the *zeitgeist*, the spirit of the times, in attempting to explain changes in psychological practices and topics over the years—see the first hundred pages of Watson and Campbell's (1963) edited collection of Boring's articles. We can do no better than to invoke a changing *zeitgeist*, even though we recognize that *zeitgeist* can just mean "something happened and it is difficult to pinpoint a specific cause."

The same year we published our article in *Psychological Science*, we published a review of the testing effect in *Perspectives in Psychological Science* (Roediger & Karpicke, 2006a). That article reviewed the history of studies on the testing effect from several disparate literatures over nearly a century. We suspect that the review also attracted interest and helped call attention to our empirical article. As of December 22, 2017, the Roediger and Karpicke (2006b) empirical article has been cited 1,589 times and the review article (Roediger & Karpicke, 2006a) has been cited 1,340 times. We suspect that the articles had a symbiotic effect.

A final point, noted above, is that our experiments produced striking testing or retrieval practice effects and showed that the effect can be pronounced after a long retention interval. The cross-over interaction in Figure 1 is unusual, and the debate in the field of

human learning and memory has usually been aimed at the issue of whether any variable can be shown to slow forgetting (Slamecka & McElree, 1983; see too Loftus, 1985). We showed that retrieval practice is one variable that slows forgetting relative to rereading of material. Perhaps this is another reason that the article has had impact, although we suspect that the practical implications were more likely the reason.

Our results validate William James's (1890) assertion about remembering that he apparently came upon by observing his own attempts to memorize:

A curious peculiarity of our memory is that things are impressed better by active than by passive repetition. I mean that in learning (by heart, for example), when one almost knows the piece, it pays better to wait and recollect by an effort from within than to look at the book again. If we recover words in the former way, we shall probably know them the next time; if in the latter way, we shall very likely need the book once more. (p. 646)

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

References

- Abbott, E. E. (1909). On the analysis of the factors of recall in the learning process. *Psychological Monographs*, *11*, 159–177.
- Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, *87*, 659–701. doi:10.3102/0034654316689306
- Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 123–144). Hillsdale, NJ: Erlbaum.
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.
- Blunt, J. R., & Karpicke, J. D. (2014). Learning with retrieval-based concept mapping. *Journal of Educational Psychology*, *106*, 849–858.
- Butler, A. C., Marsh, E. J., Slavinsky, J. P., & Baraniuk, R. G. (2014). Integrating cognitive science and technology improves learning in a STEM classroom. *Educational Psychology Review*, *26*, 331–340.
- Butler, A. C., & Roediger, H. L., III. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, *19*, 514–527.
- Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology*, *19*, 619–636.
- Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, *20*, 633–642.
- Einstein, G. O., Mullet, H. G., & Harrison, T. L. (2012). The testing effect: Illustrating a fundamental concept and changing study strategies. *Teaching of Psychology*, *39*, 190–193.
- Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior*, *10*, 562–567.
- James, W. (1890). *Principles of psychology*. New York, NY: Holt.
- Kang, S. H. K. (2017). The benefits of interleaved practice for learning. In J. C. Horvath, J. M. Lodge, & J. Hattie (Eds.), *From the laboratory to the classroom: Translating science of learning for teachers* (pp. 79–93). New York, NY: Routledge.
- Kang, S. H. K., McDermott, K. B., & Roediger, H. L., III. (2007). Test format and corrective feedback modulate the effect of testing on memory retention. *The European Journal of Cognitive Psychology*, *19*, 528–558.
- Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331*, 772–775.
- Karpicke, J. D., Blunt, J. R., & Smith, M. A. (2016). Retrieval-based learning: Positive effects of retrieval practice in elementary school children. *Frontiers in Psychology*, *7*, Article 350. doi:10.3389/fpsyg.2016.00350.
- Karpicke, J. D., Butler, A. C., & Roediger, H. L., III. (2009). Metacognitive strategies in student learning: Do students practice retrieval when they study on their own? *Memory*, *17*, 471–479.
- Karpicke, J. D., & Roediger, H. L., III. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, *57*, 151–162.
- Karpicke, J. D., & Roediger, H. L., III. (2008). The critical importance of retrieval for learning. *Science*, *319*, 966–968.
- Larsen, D. P., Butler, A. C., & Roediger, H. L., III. (2009). Repeated testing improves long-term retention relative to repeated study: A randomized controlled trial. *Medical Education*, *43*, 1174–1181.
- Loftus, G. R. (1985). Evaluating forgetting curves. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 397–406.
- Lyle, K. B., & Crawford, N. A. (2011). Retrieving essential material at the end of lectures improves performance on statistics exams. *Teaching of Psychology*, *38*, 94–97.
- 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.
- McDermott, K. B., Agarwal, P. K., D'Antonio, L., Roediger, H. L., III, & McDaniel, M. A. (2014). Both multiple-choice and short-answer quizzes enhance later exam performance in middle and high school classes. *Journal of Experimental Psychology: Applied*, *20*, 3–21.
- Putnam, A. L., Nestojko, J. F., & Roediger, H. L., III. (2017). Improving student learning: Two strategies to make it stick. In J. C. Horvath, J. Lodge, & J. A. C. Hattie (Eds.), *From the laboratory to the classroom: Translating the*

- science of learning for teachers* (pp. 94–121). New York, NY: Routledge.
- Roediger, H. L., III. (2013). Applying cognitive psychology to education: Translational educational science. *Psychological Science in the Public Interest*, *14*, 1–3.
- Roediger, H. L., III, & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*, 181–210.
- Roediger, H. L., III, & Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255.
- Roediger, H. L., III, Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. In J. Mestre & B. Ross (Eds.), *Psychology of learning and motivation: Cognition in education* (pp. 1–36). Oxford, England: Elsevier.
- Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*, 1432–1463.
- Schwieren, J., Barenberg, J., & Dutke, S. (2017). The testing effect in the psychology classroom: A meta-analytic perspective. *Psychology Learning & Teaching*, *16*, 179–196.
- Slamecka, N. J., & McElree, B. (1983). Normal forgetting of verbal lists as a function of their degree of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 384–397.
- Soderstrom, N. C., Kerr, T. K., & Bjork, R. A. (2016). The critical importance of retrieval—and spacing—for learning. *Psychological Science*, *27*, 223–230.
- Thompson, C. P., Wenger, S. K., & Bartling, C. A. (1978). How recall facilitates subsequent recall: A reappraisal. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 210–221.
- Watson, R. I., & Campbell, D. T. (1963). *Edwin G. Boring—History, psychology & science: Selected papers*. New York, NY: John Wiley & Sons.
- Wheeler, M. A., Ewers, M., & Buonanno, J. (2003). Different rates of forgetting following study versus test trials. *Memory*, *11*, 571–580.
- Wheeler, M. A., & Roediger, H. L., III. (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science*, *3*, 240–245.