

6 The Relative Benefits of Studying and Testing on Long-Term Retention

Henry L. Roediger, III and John F. Nestojko

Besides the storage that takes place during list presentation, additional storage may occur during the long-term retrieval process itself.

—Raaijmakers & Shiffrin (1981, p. 95)

In most educational situations, both teachers and students make a sharp distinction between learning and assessment. Learning occurs when students read their textbooks, listen to lectures, take notes, review their books and notes, and work in collaborative study groups. Students' learning is assessed from time to time by the requirement to take quizzes, tests, and exams. Little thought is given to the idea that the act of assessment may affect learning, except perhaps in the writing of term papers and the like. Teachers and students think tests are like dipsticks, dropped into the students' heads every so often to measure what they have learned, but without having any effect on the process of learning itself.

The same sharp distinction between studying and testing is embedded in experimental approaches to learning. Psychologists interested in learning started out with the assumption that testing measures but does not affect memory. The classic method to measure learning has been the venerable study/test method (Roediger & Arnold, 2012). Subjects in psychology experiments study a list of elements (nonsense syllables, words, pictures) and then take a test of some sort on that material (such as a free recall test). Then these study/test cycles are repeated. When performance (e.g., number of words recalled) is plotted on the ordinate against trials on the abscissa, the standard learning curve unfolds. A typical one is represented in the standard condition in Figure 6.1.

Psychologists have spent 130 years debating how and why learning occurs, since Ebbinghaus (1885/1964) began experimental study of the topic. One assumption is that every time an element (e.g., a word) is presented, its mental representation or its memory trace accrues a bit of strength; over trials, as strength grows, a threshold is reached that permits the subject (or student) to recall the information. This is the incremental assumption of learning, embodied in Clark Hull's learning theory (e.g., Hull, 1943) and many others.

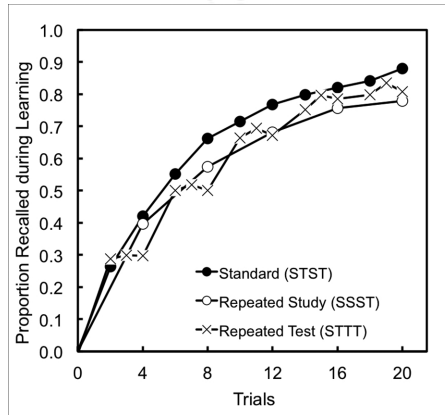


Figure 6.1 Proportion recalled on tests during learning across three learning conditions. Adapted from Karpicke and Roediger (2007; Experiment 1).

Another theory of how learning occurs is the all-or-none view of trace storage. According to the all-or-none hypothesis, trace strength grows as a step function; traces go from no strength to full strength on one trial (Guthrie, 1942). Learning appears gradual in Figure 6.1 due to an averaging artifact; over many subjects and many items, the process looks gradual, but this belies the fact that the underlying process is all-or-none. Rock (1957) and Estes (1960) both advanced this view in somewhat different formulations.

Although by the mid-1960s the incremental assumption seemed to have won the day, Roediger and Arnold (2012) recently argued that the issue has never been properly decided and that the grounds for refuting the all-or-none assumption should be re-examined. However—and the reason for bringing up this debate here—theorists in both camps shared one assumption in common: Both groups assumed that learning occurred during the study phase of the learning experiment and that the test phase existed only to measure what had been learned (and how).

The purpose of this chapter is to question this overarching assumption—which is still prevalent today in both education and experimental psychology—and present evidence that it is wrong. Richard Shiffrin and his colleagues were prescient in this matter, as the opening quote of the chapter exemplifies, by arguing that retrieval affects and changes learning as well as measuring it (see too Bjork, 1975). As we shall see, the importance of retrieval practice and tests were built into Shiffrin's and his colleagues' models.

COMPARING STUDY AND TEST TRIALS

Tulving (1967) called into question the traditional assumption about the relation of studying and testing in learning experiments by comparing the standard study/test method to other possible arrangements. In a free recall

experiment, Tulving had subjects study 36 words and compared performance across cycles of four trials with three possible arrangements: The standard study/test version (STST for alternating study and test trials); a study version with three study trials and one test in each cycle (SSST); and a test version with only one study trial and three tests (STTT). The amount of time given for study and test trials was equated, so the two types of events could be directly compared without any confounds with respect to time on task. These types of cycles—STST, SSST, and STTT—were repeated six times for 24 trials in all.

If subjects learn during study trials and tests only serve to measure performance, then the SSST cycles should lead to much better recall at the end of learning (18 study trials in total) than the standard alternating STST method (12 total study trials). Of course, both these conditions should surpass the test-dominant cycles (with only six study trials in total). However, Tulving discovered that the learning curve led to about the same eventual level of recall on the very last test trial for all three schedules of study/test practice. That is, six study trials (with 18 test trials) led to as much learning as 18 study trials (with six tests). Tulving (1967) concluded that, within the context of these conditions, learning over trials “depends primarily on the time spent on the task and that it is relatively little affected, if at all, by the distribution of this time between studying and recalling the material” (p. 181). Further, the act of recalling during a test “seems to facilitate subsequent . . . recall approximately to the same extent as does study of the material” (p. 181). When Rich Shiffrin and his collaborators developed their search models of recall, they included the assumption that “during recall, additional information is stored in long-term store about the particular items recalled” (Shiffrin, 1970, p. 404), and based this idea partly on Tulving’s results. This assumption proved to be critical for accounting for memory phenomena such as hypermnesia (e.g., Roediger & Thorpe, 1978) and others.

Karpicke and Roediger (2007) replicated Tulving’s (1967) results 40 years later and added a twist. One reason for the replication is that Tulving had permitted subjects only 36 seconds to recall (aloud) 36 words, so Karpicke and Roediger worried that such little time may not have permitted subjects to reach asymptotic levels of recall (see Roediger & Thorpe, 1978). Karpicke and Roediger essentially repeated Tulving’s experiment by having subjects study 40 words with varying patterns of study and recall trials (with study and test trials lasting 2 minutes) across three different conditions: STST, SSST, and STTT. The learning curves across five such cycles are shown in Figure 6.1. Karpicke and Roediger mostly replicated Tulving’s results, although in these experiments the standard study/test condition produced somewhat better performance than did the other two conditions. In a later experiment, Karpicke and Roediger showed that there seems to be something special about alternating study and test trials relative to the other conditions, because other arrangements (e.g., SSTT) produced worse performance, too. Having a study trial following immediately

after a test may serve as a feedback trial and lead to test-potentiated learning (e.g., Arnold & McDermott, 2013; Izawa, 1971).

The new twist in the Karpicke and Roediger (2007) experiment just described is that they tested subjects one last time a week later, providing ten minutes to recall as many items as possible in a final free recall test. How would the three schedules of learning affect retention on a delayed test? The answer is that the STST condition led to best final recall (0.68), but the STTT condition produced better recall (0.64) than the SSST condition (0.57). In short, having 15 tests (and five study trials) led to better long-term recall than 15 study trials (and five tests). Note that this advantage occurred despite the fact that on study trials all 40 words were presented whereas on test trials subjects were only re-exposed to the number of words they could recall. Considering an *intact exposure* as a study trial of a single word or as a successful recall of a single word, the average number of intact exposures of items for the SSST and STTT conditions was 728 and 568. Thus the act of recall during tests provides a much more powerful mnemonic boost for long-term retention than does repeated studying, because even though the number of intact exposures to words were many fewer for subjects in the STTT condition, they recalled more items after a week than did subjects in the SSST condition.

The outcome described in the last sentence refers to the testing effect or the retrieval practice effect (both names are used): The act of retrieval often provides a powerful benefit to recall or recognition on a delayed test relative either to no intervening activity (Wheeler & Roediger, 1992) or to restudying (Roediger & Karpicke, 2006a). This effect has been known for 100 years but only intensively studied in the past 20 (see Roediger & Karpicke, 2006b for a historical review and Roediger & Butler, 2011 or McDermott, Arnold, & Nelson, 2014 for more recent reviews).

To summarize this introductory section, a huge literature on learning and remembering has emphasized encoding factors—manipulations at study—as though the great problem of memory is encoding and storage, of getting information “into memory.” That has been the standard assumption, with the emphasis on initial learning during study in both education and in the study of learning and memory. However, the testing effect shows that practicing retrieval—practicing calling information to mind when it is needed—is also crucial. To quote Tulving (1991), “the critical process in memory is retrieval” (p. 91; see too Roediger, 2000).

A PUZZLE SOLVED

The foregoing considerations lead to a paradox: In the Tulving (1967) and Karpicke and Roediger (2007) experiments just reviewed, study events and test events seem about equally potent in producing learning on immediate tests, although testing produces larger effects than studying on delayed tests (see too Hogan & Kintsch, 1971; Roediger & Karpicke, 2006a; Wheeler,

Proof

Ewers, & Buonanno, 2003). However, many other researchers have found testing effects (especially when the tests are given with feedback) on a final criterial test given shortly after learning (e.g., Carrier & Pashler, 1992; see Roediger & Karpicke, 2006b for many other examples). These considerations lead to a puzzle: Does testing actually affect the learning curve? Tulving’s analysis would seem to suggest not, because he found that study trials and test trials can substitute for one another (within limits) with each having roughly the same effect. Yet this conclusion seems odd, given the power of retrieval practice in other experimental techniques. Another way to pose this issue is by asking: If one plotted a learning curve without cumulative tests, would it be just like the standard learning curve? If so, what might this mean, given other evidence that testing is efficacious for learning?

To answer the questions just posed, Roediger and Smith (2012) conducted three experiments using free recall and paired-associate learning of words. We report Experiment 1 here. The basic idea was to compare the learning curve derived from the standard study/test procedure across eight ST sequences to that of a “pure study” condition. In the latter, different groups of subjects studied items the same number of times (and with the same spacing) as subjects in the standard condition, but with a different group of subjects receiving either two, three, four, six, or eight study episodes before receiving a single test. The basic contrast is shown in the design outlined in Table 6.1. By using this technique, we could compare whether the standard STST learning procedure with an increasing number of tests would produce a benefit relative to the same number of study trials with no tests until the final one. Thus, subjects study the list two, three, four, six, or eight times with spaced presentations like those in the standard condition but then receive only one final test. In this way, we can sweep out a between-subjects learning curve without repeated testing.

Table 6.1 Design of Roediger and Smith (2012). The learning curve from the standard condition (top row) was compared to a learning curve derived from the other five conditions in which subjects studied lists as in the standard condition but performed a filler activity rather than taking cumulative tests. Subjects in these conditions were tested only once.

	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6	Period 7	Period 8
Standard	S T	S T	S T	S T	S T	S T	S T	S T
Study 8	S –	S –	S –	S –	S –	S –	S –	S T
Study 6	S –	S –	S –	S –	S –	S T		
Study 4	S –	S –	S –	S T				
Study 3	S –	S –	S T					
Study 2	S –	S T						

Note: S denotes a study trial, T denotes a test trial, and – denotes a filler trial.

Proof

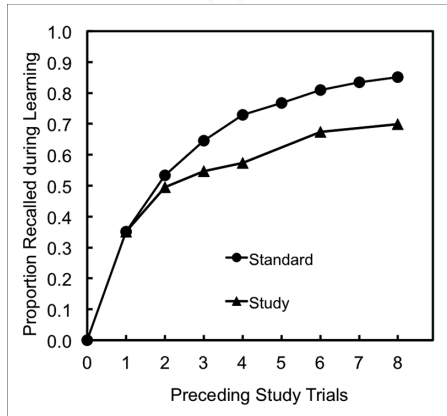


Figure 6.2 Proportion recalled on tests in either a standard learning condition (STST. . .) or a “pure study” learning condition in which subjects received the same number of study episodes of the list as in the standard condition but with only one test given. Adapted from Roediger and Smith (2012).

The results are shown in Figure 6.2, where it can be seen that the standard study/test procedure produced faster and greater learning than the “pure study” procedure. Thus, testing *does* affect the learning curve. Roediger and Smith’s (2012) Experiment 2 replicated this outcome with paired-associate learning, and Experiment 3 showed it was not merely due to more exposures on the tests producing additional learning, because adding more study trials (in place of tests) in the pure study condition did not overcome the advantage of testing in the standard study/test condition. In short, testing clearly matters in the development of learning. Given that learning has been studied for 130 years, it is surprising that no one asked this question until recently. Test trials clearly matter to learning, but what are their relative contributions to long-term memory? We turn to this issue next.

PITTING STUDY TRIALS AGAINST TEST TRIALS FOR LONG-TERM RETENTION

Zaromb and Roediger (2010) conducted experiments designed to shed light on the mechanism of the testing effect in free recall, of why tests often produce better performance than an equivalent number of study opportunities. We describe their experiments’ purpose (and findings) first and then their relevance for this chapter. Zaromb and Roediger used categorized lists (e.g., types of birds, articles of furniture) and presented items in random order to subjects in various sequences of study and test trials. They were interested in the effects of test trials relative to study trials on subjects’ organization of

the lists, hypothesizing that test trials may lead to greater recall than study trials because they permit subjects to develop organizational schemas (or retrieval plans) for recall. They found evidence in support of this hypothesis by measuring both categorical clustering in recall (the tendency to recall items together from the same categories; Bousfield, 1953) and especially in subjective organization (the tendency to recall words in the same sequence from one test trial to another; Tulving, 1962). They showed that tests led to greater organization in final recall relative to study trials, confirming the hypothesis that the power of tests may arise in part because subjects develop plans for retrieving information. These results are fully compatible with Raaijmakers and Shiffrin's (1980, 1981) Search of Associative Memory (SAM) model, which postulated that recalled items become increasingly associated with one another during the process of retrieval and hence can guide retrieval better on future recall attempts.

Although Zaromb and Roediger (2010) set out to investigate the effect of retrieval on organization of materials, their Experiment 1 addressed a question directly relevant to this chapter: What is the effect of repeated studying versus repeated testing on long-term retention? In their Experiment 1, subjects studied and/or were tested on a list of 50 words for a total of eight times. However, the proportion of study and test trials traded off across three conditions. In one condition, subjects received alternating study and test trials in the standard method (ST ST ST ST), with four study trials and four free recall test trials. In a second condition, subjects studied the list six times and took two free recall tests (ST SS ST SS), and in a third condition subjects studied the list eight times (SS SS SS SS). Subjects took a final free recall test two days after learning.

Note that across these conditions, the number of study trials increased from four to six to eight. According to Hintzman (1976), "the fact that repetition improves memory, established empirically in Ebbinghaus's experiments and in thousands of studies since, seems beyond dispute" (p. 47). If this statement is true—and most psychologists would likely agree that it is—then subjects in the condition with the greatest number of study trials (i.e., the most repetitions) should recall the most. The results displayed in Figure 6.3 tell a different story: Taking the data at face value, there is a negative effect of repeated studying on a test two days after learning. Eight study trials produced the worst free recall (0.17), six study trials were somewhat better (0.25), and four study trials were the best (0.39), contradicting years of research showing that repetition improves memory. Looked at another way, however, the results are more sensible: Final test performance increased with the number of test trials administered during learning (zero, two, and four tests produced 0.17, 0.25, and 0.39 proportion correct recall, respectively, two days later). Thus the conclusion from this analysis of Zaromb and Roediger (2010, Experiment 1) is that the number of test trials, rather than the number of study trials, determined long-term retention. The more test trials during the initial learning phase, the better was recall

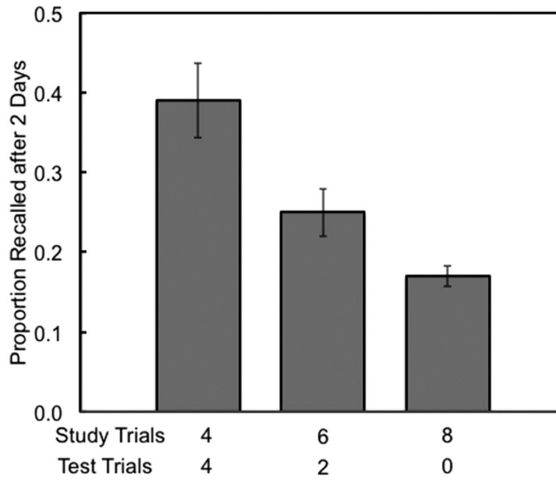


Figure 6.3 Proportion recalled on the final test two days after learning. As indicated, means are plotted as a function of increasing study trials and decreasing test trials. Error bars represent 95% confidence intervals. Adapted from Zaromb and Roediger (2010; Experiment 1).

on the final test (despite the fact that the conditions with fewer test trials had more study opportunities). These results show (again) that testing has power beyond simple item repetition. The condition with eight study trials involved 400 intact exposures relative to 342 for the six study, two test trial condition and 301 for the four study, four test trial condition.

There is a fly in the ointment in this analysis, however, with regard to the relative effect of study and test trials on long-term retention. Astute readers will have noticed that the conclusion is on somewhat shaky ground because the number of study trials and test trials was confounded in this experiment, as the abscissa of Figure 6.3 makes clear. As the number of test trials increased, the number of study trials decreased (and vice versa). Therefore, no strong conclusions can be drawn about the relative benefits of study and test, although certainly the conclusion that test events are more potent memory enhancers than study events agrees with other research (e.g., Karpicke & Roediger, 2007, 2008).

Nestojko and Roediger (2014) set out to correct this confounded state of affairs in a new set of experiments by orthogonally manipulating the number of study trials and test trials in the learning phase of an experiment to analyze how various combinations would affect recall two days later. We employed the same basic paradigm developed by Zaromb and Roediger (2010) in that we had different groups of subjects study a 50-word list and then take free recall tests, with a final test two days later. However, we crossed study trial repetitions (one, two, or four) with test trial repetitions (zero, one, two, or

four) in order to determine the relative power of study and test events when the two types were not confounded as in Zaromb and Roediger's Experiment 1. (Our manipulation was not strictly orthogonal, because we included a condition with zero test trials and it makes no sense to have a condition with zero study trials). Our design produced twelve between-subjects conditions instantiated in the learning phase on the first day of the experiment: One, two, or four study trials with zero tests (S, SS, SSSS); the same study conditions with one test (ST, SST, SSSST); the same study conditions with two tests (STT, STST, SSTSST); and finally the same study conditions with four tests (STTTT, STTSTT, STSTSTST). A different group of subjects (from Amazon's Mechanical Turk recruitment site) was assigned to each condition and all were tested two days later.

We consider performance during initial learning first and then performance on the final test after 48 hours. Consistent with Tulving's (1967) results, Nestojko and Roediger (2014) found that study and test trial repetitions had similar positive effects on learning, as measured by performance on tests administered during acquisition. We do not present those results here, but they were generally consistent with those of Tulving (1967) and Karpicke and Roediger (2007; presented in Figure 6.1), who also did not find an advantage for testing conditions during learning.

The critical results are those from the final test given two days after learning, which are shown in Figure 6.4. The light bars in the figure depict performance in the one, two, and four study trial conditions when the data are

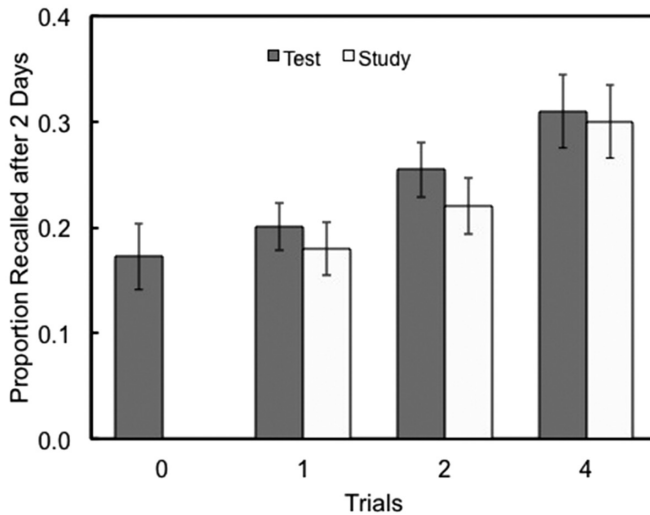


Figure 6.4 Proportion recalled on the final test two days after learning (Nestojko & Roediger, 2014). Test trial data are collapsed across study trials, and study trial data are collapsed across test trials. Error bars represent standard error of the mean.

collapsed across test trials in order to determine the effects of study trials on long-term recall. Not surprisingly, repeated study trials did not impair long-term retention (as the analysis in Figure 6.3 might imply), but rather increasing study trials enhanced delayed recall. Thus, the seemingly curious result reported by Zaromb and Roediger (2010) of study repetitions seeming to hurt long-term recall was due to the confounding with testing. Hintzman's (1976) quote cited earlier remains on-target and the psychological world is normal.

The apparent surprise in our data comes from the other analysis represented in the dark bars of Figure 6.4. These data show the effects of repeated tests when the data are combined across the various study conditions. Repeated testing had a positive effect on delayed recall, but the effect was no greater than that for the repeated study conditions. (An analysis of variance showed that both number of study trials and number of test trials had significant effects on long-term recall with no interaction). The fact that the positive effect of repetition was approximately the same whether repetitions during learning were study trials or test trials is reminiscent of Tulving's (1967) conclusions from immediate free recall, but is unlike the findings of many researchers who examined effects of study trials and test trials on delayed recall (Karpicke & Roediger, 2007, 2008; Roediger & Karpicke, 2006a, among others).

Thus the Nestojko and Roediger (2014) results lead to yet another puzzle for consideration in this chapter, but one that we think is understandable. Although we found that repeated testing has roughly the same effects as repeated studying on long-term retention—quite different from other results we have reviewed—the design of our experiment may help to explain why repeated tests did not appear to enhance long-term retention more than repeated study trials. Consider the disparate power of the two manipulations. On every study trial, subjects studied all 50 words (or 100%). However, on repeated test trials, subjects only experienced the items they could recall on the initial tests. This number averaged about 18.5 words (or 37%) across the initial test trials. Thus, one way of looking at our data is that retrieving only 37% of the items is just as powerful as studying 100% of them. Hence, the power of testing still shines through in our results when the data are viewed in this light. Even though subjects failed to recall 63% of the items on the initial tests, test trials were as powerful as study trials on long-term retention. If the tests had been given under conditions in which recall was greater (e.g., lists of 15 words rather than 50, perhaps), the power of testing may have outweighed the effects of study trials. The point is that items that were not recalled in free recall do not produce a testing effect (unless feedback is given). In fact, many years ago Shiffrin and his colleagues made just this point when they wrote, “in a recall situation. . . it seems likely that the major increase in strength would occur after successful recall takes place, with little or no increase after a failure to recall” (Raaijmakers & Shiffrin, 1981, p. 128).

Karpicke and Roediger (2008) provided a more powerful way to compare the relative benefit of study and test trials. In a paired-associate learning experiment, subjects were given study-test-feedback practice until they successfully recalled a word pair. After the pair was recalled, it was given either repeated study practice or repeated test practice. The outcome in this procedure (which does not confound studying and testing during learning) showed that repeated tests had a large effect on recall two days later whereas as repeated study trials had little to no effect. Thus, in a procedure in which study and test trials can be more carefully controlled, retrieval practice appears much stronger than repeated studying.

CONCLUSION

The relative balance of study and test opportunities in facilitating long-term retention is an important issue for the theoretical and empirical study of learning and also has important educational implications. Given this consideration, we note in closing the curious fact that learning has been studied empirically for 130 years and yet only a handful of experiments have dealt with the fundamental question of the relative roles that study and test events play in the process. To return to the theme with which we began the chapter, we can surmise that the reason for this neglect is that all researchers implicitly or explicitly assumed that testing has no effect on learning. Richard Shiffrin's writings from over the past half-century show that he was aware of how retrieval would affect learning, and our experiments reported in this chapter confirm his assumptions. As Shiffrin and Raaijmakers (1992) wrote, "when a successful recovery does occur, it is assumed that the strengths between each cue and the image [memory trace] in question are increased" (p. 72). Given the results published in this chapter, this is no longer an assumption, but a well-verified empirical result.

AUTHOR NOTES

We thank Pooja Agarwal, Andy DeSoto, Adam Putnam and Jeroen Raaijmakers for comments on an earlier version of this chapter. The research reported was supported by a grant from the James S. McDonnell Foundation.

REFERENCES

- Arnold, K. M., & McDermott, K. B. (2013). Test-potentiated learning: Distinguishing between direct and indirect effects of tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39, 940–945.

Proof

- 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.
- Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, 49, 229–240.
- Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, 20, 633–642.
- Ebbinghaus, H. (1885/1964). *Memory: A Contribution to Experimental Psychology* (Translated by H. A. Ruger & G. E. Bussenius, 1913). New York, NY: Dover.
- Estes, W. K. (1960). Learning theory and “the new mental chemistry.” *Psychological Review*, 67, 207–223.
- Guthrie, E. R. (1942). Conditioning: A theory of learning in terms of stimulus, response, and association. *Yearbook of the National Society of Education*, 41, 17–60.
- Hintzman, D. L. (1976). Repetition and memory. In G. Bower (Ed.), *The Psychology of Learning and Motivation* (pp. 47–91). New York, NY: Academic Press.
- Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term retention and recall. *Journal of Verbal Learning and Verbal Behavior*, 10, 562–567.
- Hull, C. L. (1943). *Principles of Behavior: An Introduction to Behavior Theory*. Oxford, England: Appleton-Century.
- Izawa, C. (1971). The test trial potentiating model. *Journal of Mathematical Psychology*, 8, 200–224.
- Karpicke, J. D., & Roediger, H. L. (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. (2008). The critical importance of retrieval for learning. *Science*, 319, 966–968.
- McDermott, K. B., Arnold, K. M., & Nelson, S. M. (2014). The testing effect. In T. Perfect and S. Lindsay (Eds.), *Sage Handbook of Applied Memory* (pp. 183–200). Thousand Oaks, CA: Sage.
- Nestojko, J. F., & Roediger, H. L. (in preparation). The relative power of study and test trials in enhancing long-term retention.
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory*, Vol. 14 (pp. 207–262). New York, NY: Academic Press.
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, 88, 93–134.
- Rock, I. (1957). The role of repetition in associative learning. *American Journal of Psychology*, 70, 186–193.
- Roediger, H. L. (2000). Why retrieval is the key process to understanding human memory. In E. Tulving (Ed.), *Memory, Consciousness, and the Brain: The Tallinn Conference* (pp. 52–75). Philadelphia, PA: Psychology Press.
- Roediger, H. L., & Arnold, K. M. (2012). The one-trial learning controversy and its aftermath: Remembering Rock (1957). *American Journal of Psychology*, 125, 127–143.
- Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, 15, 20–27.
- Roediger, H. L., & Karpicke, J. D. (2006a). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17, 249–255.
- Roediger, H. L., & Karpicke, J. D. (2006b). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1, 181–210.

- Roediger, H. L., & Smith, M. A. (2012). The “pure-study” learning curve: The learning curve without cumulative testing. *Memory & Cognition*, *40*, 989–1002.
- Roediger, H. L., & Thorpe, L. A. (1978). The role of recall time in producing hypermnesia. *Memory & Cognition*, *6*, 296–305.
- Shiffrin, R. M. (1970). Memory search. In D. A. Norman (Ed.), *Models of Memory* (pp. 375–447). New York, NY: Academic Press.
- Shiffrin, R. M., & Raaijmakers, J. G. W. (1992). The SAM retrieval model: A retrospective and prospective. In A. Healy, S. Kosslyn, & R. M. Shiffrin (Eds.), *From Learning Processes to Cognitive Processes: Essays in Honor of William K. Estes, Vol. 2* (pp. 69–86). Hillsdale, NJ: Erlbaum.
- Tulving, E. (1962). Subjective organization in the free recall of “unrelated” words. *Psychological Review*, *69*, 344–354.
- Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, *6*, 175–184.
- Tulving, E. (1991). Interview with Endel Tulving. *Journal of Cognitive Neuroscience*, *3*, 89–94.
- Wheeler, M. A., Ewers, M., & Buonanno, J. F. (2003). Different rates of forgetting following study versus test trials. *Memory*, *6*, 571–580.
- Wheeler, M. A., & Roediger, H. L. (1992). Disparate effects of repeated testing: Reconciling Ballard’s (1913) and Bartlett’s (1932) results. *Psychological Science*, *3*, 240–245.
- Zaromb, F. M., & Roediger, H. L. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition*, *38*, 995–1008.

Taylor & Francis
Not for distribution

Proof