

The Indirect Effects of Testing: Can Poor Performance in a Vocabulary Quiz Lead to Long-Term L2 Vocabulary Retention?

Kohei Kanayama^a and Kiwamu Kasahara^b

^a*Sapporo Otani High School*; ^b*Hokkaido University of Education*

doi: <https://doi.org/10.7820/vli.v07.1.kanayama.kasahara>

Abstract

Taking a test on learned items enhances long-term retention of these items. However, it is believed that good performance in a test contributes to subsequent high retention of the tested items while poor performance does not. Recent studies have sought to find the optimal way to make up for this poor performance, and have indicated that giving the subsequent learning session soon after the test is one such way. This study is different from previous studies in that we used L1–L2 word pairs to examine whether restudying immediately after the failure in the test is useful for long-term retention. First, in the initial study session, all the participants ($n = 52$) were shown and asked to remember 20 English and Japanese word pairs (e.g., deceit:詐欺). A week later, Group A took the first test session (Initial Test) before the restudy session. On the contrary, Group B took the restudy session before the Initial Test. An hour after this session, both groups took Posttest 1. Then, Posttest 2 was conducted a week after Posttest 1. The results showed that Group A had significantly lower scores than Group B in the Initial Test (2% vs. 55%). However, the results were reversed in Posttest 1 (84.2% vs. 53.2%) and Posttest 2 (55% vs. 43.5%). This study found that a restudy session soon after poor performance in the Initial Test enhanced long-term L2 vocabulary retention because learners benefited from the indirect effects of testing. Thus, English teachers should take such effects into consideration when organizing vocabulary quizzes and restudy sessions.

1 Introduction

There is a great deal of evidence that taking a test on learned items can strengthen long-term retention of these items. This phenomenon has been widely demonstrated as the *testing effect* (Barcroft, 2007; Carpenter, Pashler, & Vul, 2006; Halamish & Bjork, 2011; Karpicke & Roediger, 2007; Meyer & Logan, 2013). Previous studies have demonstrated that tests have positive effects on any educationally relevant materials such as fact learning (Agarwal, Finley, Rose, & Roediger, 2017; Butler & Roediger, 2007; Carpenter, Pashler, Wixted, & Vul, 2008; Roediger, Agarwal, McDaniel, & McDermott, 2011a), learning from reading passages (McDaniel, Anderson, Derbish, & Morrisette, 2007a; Roediger & Karpicke, 2006b), picture-L2 word pairs (Barcroft, 2007), cue–target word pair learning (Carpenter et al., 2006; Cull, 2000; Halamish & Bjork, 2011; Karpicke & Roediger,

2007; Roediger & Marsh, 2005), and L1–L2 pair learning (Kanayama & Kasahara, 2015). Thus, a test itself can be a powerful tool to improve students' performance in a later test. However, the testing effect has been ignored in actual school settings (Carpenter et al., 2008; Karpicke & Grimaldi, 2012; McDaniel et al., 2007). In other words, teachers have generally viewed tests as just tools to evaluate students' current learning achievements (Butler & Roediger, 2007), to assign grades to the students (Roediger, Putnam, & Smith, 2011b; Vojdanoska, Cranney, & Newell, 2010), or just as means to assess what has been learned (Blunt & Karpicke, 2014; Karpicke & Roediger, 2007). Few teachers have seen the test as a tool to improve students' performance. However, some researchers emphasize on the importance of the test by stating that “tests should be used frequently in educational contexts not merely to assess learning, as is the standard practice, but to promote it” (Carpenter et al., 2008, p. 438), and “testing to enhance learning should be seriously considered in pedagogical theory and practice” (McDaniel et al., 2007, p. 510). Therefore, the importance of the tests as a tool to enhance students' long-term retention should be reviewed.

2 Literature Review

2.1 *The Direct Effects of Testing on Long-Term Retention*

Tests have a positive effect on learning. Roediger et al. (2011) distinguish two types of testing effect: direct and indirect effects of testing. The direct effect means that taking a test itself has a positive effect on learning (Carrier & Pashler, 1992; Roediger & Karpicke, 2006a). Learners can benefit greatly from just taking a test on learned items. One of the reasons is that tests always include *retrieval practice* (Nation, 2013). This is a process in which learners try to search for previously learned items in their memory. In case of vocabulary meaning retrieval, test-takers are often asked to retrieve L1 equivalents of target words by seeing the L2 forms of the targets. In the case of vocabulary form retrieval, they are often asked to retrieve L2 forms by seeing their L1 equivalents. The act of retrieving the items itself enhances the long-term retention of them (Karpicke, 2012) because it requires the learners to make a mental effort to take out something from their memory with the help of cues, which can strengthen connections between the cues and the targets. These reinforced connections, or deeper traces in memory, can contribute to long-term retention (Craik & Tulving, 1975). If the items are successfully retrieved from memory, they will be established more strongly in learners' minds. As a result, the learners can put the successfully retrieved items into their long-term memory (Tullis, Finley, & Benjamin, 2013).

Halamish and Bjork (2011) found a positive effect of retrieval practice on vocabulary learning. In their study, a retrieval group and a non-retrieval group were presented with L1 cue–target word pairs in random order for 3 seconds each (e.g., *RENT: HOUSE*). After the first presentation, the non-retrieval group restudied each word pair for 6 seconds in the other two cycles. On the contrary, the retrieval group looked at each cue word and a part of the target word for 6 seconds (e.g., *RENT: __SE*) in the remaining two cycles. An immediate test asking both groups to recall the targets for the cue words (e.g., *RENT: ____?*) found that the retrieval

group recalled more of the target words than the non-retrieval group (39% vs. 13%). Kanayama and Kasahara (2015) supported the findings of Halamish and Bjork, investigating the effects of word retrieval on L1–L2 pair learning (e.g., *ligament*: 靭帯). The retrieval group, who had a chance to retrieve the Japanese meanings of the target English words (e.g., *ligament*: _____?), outperformed the non-retrieval group in a test conducted a week later (31.2% vs. 22.1%).

The direct effects of testing can also be seen in passage recall. For instance, in a study by Roediger and Karpicke (2006b, Experiment 1), the researchers compared a group that took a test after the first study session with a group that had the restudy session after the first study session. The former and latter groups were called the ST and the SS groups, respectively (S stands for one study session and T stands for one test trial). The study session asked the participants to read the target English passages. In the test trial, participants were given a blank sheet of paper and asked to retrieve as much information from the passages as possible, but no correct answers were given after the test. The results showed that the ST group was superior in 1-week retention to the SS group (56% vs. 42%).

Roediger and Karpicke (2006b, Experiment 2) replicated their first experiment. The procedure was the same as for Experiment 1, except that they compared the SSSS, the SSST, and the STTT groups. The SSSS group was asked to take a study session four times in a row. Meanwhile, The SSST took a test session after three consecutive study sessions. The STTT group took a study session and three test sessions consecutively. As with Experiment 1, no feedback was given after the test. The delayed test conducted a week later found that the STTT group had a better score than the SSST group, and that the SSST group outperformed the SSSS group (61% vs. 56% vs. 40%). Roediger and Karpicke found that the more retrieval sessions the participants had during the learning phases, the better retention of the learned items they showed. These studies (Halamish & Bjork, 2011; Kanayama & Kasahara, 2015; Roediger & Karpicke, 2006b) demonstrated that learners can benefit greatly from just retrieving the learned information from their memory.

2.2 The Indirect Effects of Testing on Learning

The indirect effects of testing are another type of benefits which learners can gain through tests (Roediger & Karpicke, 2006a): test results can help test-takers to focus on items they have not mastered yet. Tests inform the students of what items they have or have not acquired (Brame & Biel, 2015; Carpenter et al., 2008; Roediger et al., 2011). Based on this experience, when they have a restudy session after the test, they are able to spend more time remembering the items which they have not mastered rather than those they have already acquired (Karpicke & Grimaldi, 2012; Soderstrom & Bjork, 2014; Son & Kornell, 2008).

Moreover, tests can prevent the learners from being overconfident (Roediger et al., 2011). Learners who have not taken any tests tend to become more confident about a later test (Brown, Roediger, & McDaniel, 2014). On the contrary, those who have taken test sessions are able to evaluate their actual learning achievement properly. The second experiment presented in Roediger and Karpicke (2006b) provides some evidence for this. After all the learning sessions, all the participants were asked to rate how well they could retrieve the learned information from the

passage they read, using a 7-point scale (1 = not very well; 7 = very well). The results revealed that the SSSS group gave higher scores ($M = 4.8$) than the SSST group ($M = 4.2$). The least confident group was the STTT group ($M = 4.0$). A week later, all the participants returned to the laboratory and took the delayed test. As mentioned above, the actual score in the delayed test showed that the STTT group was better than the SSST and the SSSS groups (61% vs. 56% vs. 40%). Figure 1 presents the participants' judgment of their learning and the actual scores of each group. The participants in the SSSS group felt that they had already achieved their learning goal. Hence, they may not have tried hard in the remaining study sessions. On the contrary, the tests prevented the participants in the STTT group from being overconfident because they understood that they had not achieved their learning goal yet. This is reliable evidence that the SSSS group, which had multiple study sessions, became overconfident about their future performance, and that the STTT group, which took the test trials, was able to avoid becoming overconfident.

In addition, tests include a beneficial by-product of indirect effects of testing. That is, if students take a test frequently, they can acquire a habit of preparing for tests. As a result, they tend to study more often and spontaneously (Putnam, Nestojko, & Roediger, 2016). Moreover, tests can help teachers allocate enough time to teach and review what most students do not understand (Roediger et al., 2011). In short, teachers as well as their students can understand actual learning achievements through tests.

A study by Karpicke and Roediger (2007) revealed that learners who gained benefits from the indirect effects of testing showed better recall performance on a delayed test than those who did not benefit from the indirect effects. In their Experiment 1, they had their participants learn cue–target word pairs in standard (STST), repeated study (SSST), or repeated test (STTT) groups. Again, no feedback (correct answers) were given after the test. The STTT group was expected to show superiority in long-term retention compared with other groups because the long-term benefits increased with the number of retrieval sessions given (Roediger & Karpicke, 2006b). Nevertheless, the test conducted at the end of all the four learning sessions found that the STST group showed superior to the STTT and the SSST (88% vs. 81% vs. 78%). Moreover, the delayed test carried out a week later showed that the STST group had a better score than the STTT and the SSST groups (68% vs. 64% vs. 57%). In addition, their Experiment 2 also compared

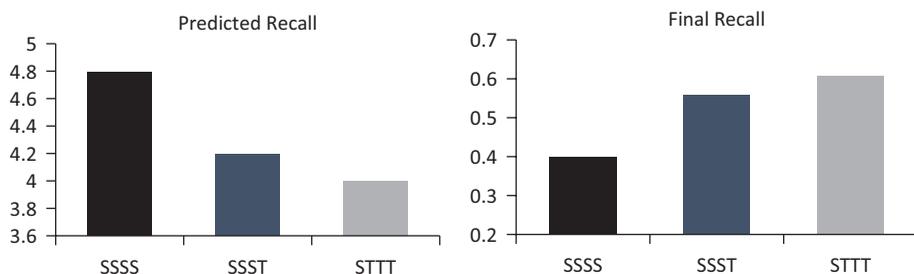


Figure 1. Predicted Scores and Final Performance on a 1-Week Delayed Test of the SSSS, SSST, and STTT Groups. Data Are Adapted From Roediger and Karpicke (2006b).

participants who learned in the STST with the SSTT groups. The results replicated their Experiment 1; the STST group showed higher retention rates than the SSTT group a week later (44% vs. 36%), indicating that the indirect effects of testing have long-term benefits.

A conceivable reason why the STST group showed better performance than the other groups (SSTT, SSST, and STTT) is that only the STST group gained more benefits from the indirect effects of testing because they had a chance to take a restudy session after a test. The first test helped the participants in the STST group notice what items they had not acquired, and helped them to evaluate themselves better. Based on this experience, they focused on the words which they had not mastered in the restudy session. Finally, they successfully recalled the items in the second test session. They benefited from both the direct and indirect effects of testing. On the contrary, the other three groups (SSTT, SSST, and STTT) benefited from the direct effects of testing (retrieval effect), but did not benefit from the indirect effects of testing because the three groups had no chance to take any restudy sessions after the tests.

Other past studies have also demonstrated that the indirect effects of testing exist on some educationally relevant materials such as learning from reading passages (Richland, Kornell, & Kao, 2009), cue–target word learning (Hays, Kornell, & Bjork, 2013; Kornell, Hays, & Bjork, 2009), and L1–L2 pair learning (Arnold & McDermott, 2013).

2.3 Previous Studies on Effects of Poor Performance in a Test on Learning

One of the reasons why indirect effects exist is that learners can narrow the items they should focus on in the subsequent study session after the test. However, it is still unclear whether the learners could gain much benefit from the indirect effects of testing even after they answered most items incorrectly. If most of the items were not recalled, such narrowing may be impossible. Therefore, it is worth investigating whether indirect effects exist even in the case where learners had a score very close to zero.

Some studies have tried to examine whether learners can benefit from the indirect effects of testing after poor performance in a test. However, these studies had some limitations. One of these studies was by Richland et al. (2009, Experiment 4). In their study, participants took the first test session in 2 minutes, where they were asked to answer the five questions relating to a short passage. Because they had not read the passage before answering the five questions, they answered nothing. After they experienced this failure, the participants took a study session in which they were asked to read a short essay (about cerebral achromatopsia) in 8 minutes. A week later, a posttest was conducted, where they were asked to answer 10 questions about the essay they read. Five out of the ten questions were in the first test session. The results showed that the five tested items were remembered better than the five untested items (55% vs. 42%). Richland et al. concluded that the learners were able to benefit from the indirect effects of testing after they showed poor performance in the previous test.

Kornell et al. (2009, Experiment 5) also had the same limitations as the study of Richland et al. In the study of Kornell et al., they had their participants remember cue–target word pairs (e.g., *train: caboose*). Participants in the read-only condition were given each pair for 13 seconds. On the contrary, those in the test condition were shown a cue word alone at first (e.g., *train: _____?*) for 8 seconds, followed by the cue and target for 5 seconds (e.g., *train: caboose*). Because the participants in the test condition had learned no word pairs, they could answer nothing in the test session. A day after the session, all the participants took the posttest, in which they were asked to recall the correct answer (e.g., *train: _____?*). It was found that the test group had a better score than the read-only condition (47% vs. 35%).

The limitation of the two studies is that there were some doubts that the participants truly showed poor performance in the test. Poor performance in a test refers to a situation in which learners failed to recall most of the learned items on tests although they had learned them before. In this respect, the participants in Richlands' and Kornells' study did not experience poor performance because they had learned no target items before the first study session. Rather, the participants in both studies benefited from the *pretesting effect*, the idea that testing before studying helps learners focus on what will be tested later. In summary, there have been some studies which demonstrated the indirect effects of testing on long-term retention, but no previous studies have investigated whether learners are able to benefit from the indirect effects of testing after their poor performance in a test.

2.4 Purposes of This Study

In order to deal with the issue mentioned above, this study compared the effects of the S-ST and S-TS conditions. (The hyphen means an interval of 1 week between the first learning and the first review session.) Learners who take the first test trial a week after the first study session are likely to forget almost all of the learned words. It is possible that this unsuccessful experience prevents them from gaining benefits from the indirect effects. Therefore, it is worth investigating whether the indirect effects of testing can function well after learners cannot recall most of the learned words. By examining the effect of the S-TS and the S-ST conditions, this study became original in that it was able to investigate whether the indirect effects of testing are useful in L1–L2 pair learning even after the learners have experienced poor performance in a test. The research question of this study is as follows:

RQ. Can the indirect effects of testing after poor performance in a test promote vocabulary learning?

3 Method

3.1 Participants

The participants in this study were 85 first-year Japanese university EFL students, and they were divided into Group A (S-TS) and Group B (S-ST).

The data from 33 participants who did not take part in one or more sessions were excluded from the analyses. Thus, the following analyses were based on 52 participants. Group A consisted of 23 students, whereas there were 29 students in Group B. All the participants had studied English in Japan for at least 6 years.

3.2 Materials

This study used 20 low-frequency English words and their Japanese equivalent pairs. These English words comprised 10 nouns and 10 verbs, which are not listed in *JACET 8000* (JACET, 2003). *JACET 8000* includes about 8000 basic English words in a frequency order which Japanese learners of English are recommended to remember. We assumed that none of the words would be known by all the participants. Each English target word and its Japanese translation are shown in Table 1.

3.3 Procedure

All the study sessions and the tests were conducted in English lessons regularly held at the university. The participants were assigned to one of the two learning conditions: the S-TS or the S-ST groups. First, both groups were asked to remember the 20 target word pairs, which were displayed on PowerPoint slides. Each word pair was presented on a screen in front of the participants. Each group had opportunities to encounter each word pair three times. In the first cycle, the participants looked at a target word and its L1 translation at the same time for 6 seconds (e.g., ligament:靭帯). In the remaining two cycles, each word pair was displayed in the same way as the first cycle for 4 seconds. The total learning time was 4 minutes and 40 seconds for both groups.

A week later, the S-TS group took the Initial Test. Immediately after the Initial Test, they took the restudying session in the same way as the first one was conducted. On the contrary, the S-ST group took the restudying session before the Initial Test. Both groups were given Posttest 1 an hour after their respective last sessions. The participants took Posttest 2 a week after Posttest 1. All the tests

Table 1. Target Nouns and Verbs

English	Japanese	English	Japanese
mutineer	反逆者	gnaw	～をかじる
lemur	キツネザル	smuggle	～を密輸する
ligament	靭帯	sterilize	～を消毒する
encroachment	侵略	sham	(病気などの)ふりをする
adhesive	接着剤	impute	人のせいにする
ointment	化粧用クリーム	belittle	～をけなす
deceit	詐欺	assail	激しく非難する
palliative	緩和剤	contort	～をねじる
janitor	用務員	foray	襲撃する
knack	こつ	immerse	(液体などに)ひたす

(Initial Test, Posttest 1, and Posttest 2) required both groups to recall all the 20 Japanese meanings for the English target words.

3.4 Scoring

Two points were awarded for each correct answer. When a student gave a correct answer but in a wrong part of speech, only one point was given. For example, in the case of the word “sterilize,” the correct answer was *sakkin-suru*, whereas one point was given for *sakkin*. The maximum score for each test was 40 points (2 points \times 20 words).

4 Results

Table 2 presents the mean scores and standard deviations of each test for Groups A and B. Figure 2 shows the mean scores in each test for Groups A and B. An analysis through a 2 (Group: S-TS and S-ST) \times 3 (Test: Initial Test, Posttest 1, and Posttest 2) mixed ANOVA revealed that there was a significant interaction between the two factors, Group and Test, $F(1, 50) = 263.8, p < 0.001, \eta^2 = 0.84$ (large effect size). In addition, a significant main effect of Test was also observed, $F(1, 50) = 223.4, p < 0.001, \eta^2 = 0.81$ (large effect size). However, there was no significant main effect of Group, $F(1, 50) = 0.53, p = 0.467, \eta^2 = 0.01$.

Table 2. Mean Scores and Standard Deviations of Each Test for Both Groups A and B (Full Mark = 40)

	Group A (S-TS)		Group B (S-ST)	
	Mean	SD	Mean	SD
Initial Test	0.82	1.85	22.0	8.78
Posttest 1	33.7	6.24	21.3	8.54
Posttest 2	22.0	8.24	17.4	7.89

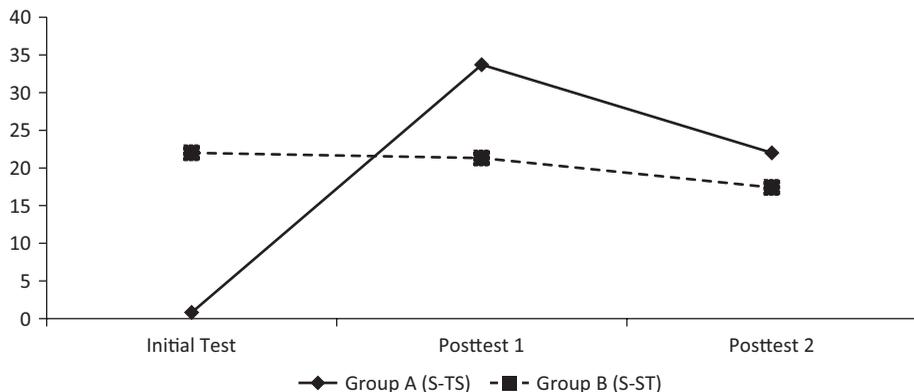


Figure 2. The Mean Scores of Each Test in Groups A and B.

Furthermore, there was a significant simple main effect between the two groups in the Initial Test, $F(1, 50) = 99.1, p < 0.001$; in Posttest 1, $F(1, 50) = 33.7, p < 0.001$; and in Posttest 2, $F(1, 50) = 4.6, p = 0.033$. This means that the S-ST group had a better score than the S-TS in the Initial Test (55% vs. 2%). However, the S-TS group showed better retention rates than the S-ST group both in Posttest 1 (84.2% vs. 53.2%), and in Posttest 2 (55% vs. 43.5%).

5 Discussion

The S-ST group showed better scores than the S-TS group in the Initial Test (55% vs. 2%). Though the S-TS group took only one study session before the Initial Test, the S-ST group took two study sessions before the Initial Test. Thus, the total amount of study time of the S-ST group (9 minutes and 20 seconds) doubled that of the S-TS group (4 minutes and 40 seconds). In addition, the S-TS group had taken the Initial Test a week after the first study session. On the contrary, the S-ST group took the restudy session soon before the Initial Test. Thus, a wide gap in scores between the S-TS and the S-ST group in the Initial Test was observed.

However, this results were reversed in Posttest 1 and Posttest 2 because the S-TS group had a significantly better score than the S-ST group both in Posttest 1 (84.2% vs. 53.2%) and in Posttest 2 (55% vs. 43.5%). As previous studies (Arnold & McDermott, 2013; Kornell, 2009) have found, the results of Posttest 1 and 2 revealed that the indirect effects of testing can be applied to L1–L2 pair learning.

The data suggested that there were indirect effects of testing for the items learners failed to recall. Indeed, 17 out of 23 participants in the S-TS group recalled nothing in the Initial Test. The average score was 0.82 out of 40 points (2%). Nevertheless, this result was reversed in Posttests 1 and 2. This suggested that learners in the S-TS group were able to benefit from the indirect effects of testing even though they had a poor performance in the Initial Test.

There are a number of reasons why the S-TS group was more successful in long-term retention of the meanings of the target items than the S-ST group. The participants in the S-TS group took the Initial Test, but they forgot most of the target words because there was a 1-week interval between the first learning session and the Initial Test. Therefore, the average scores were less than 1 point ($M = 0.82$ out of 40). They failed to retrieve the learned items from their memory in the Initial Test. However, the test may have prevented the learners in the S-TS group from being overconfident and helped them evaluate their learning achievement objectively. This should have led them to apply a great amount of effort in the restudy session. They were able to notice what words they had not recalled in the Initial Test, and were able to focus on these words during the restudy session. On the contrary, the participants in the S-ST group were not able to make the most of the indirect effects of testing because they did not have any opportunities to take the subsequent learning session after the Initial Test. They were not given a chance to make use of the information they had obtained from the test. This is likely to be the main reason why the S-TS group outperformed the S-ST group in Posttests 1 and 2.

Some limitations of this study should be noted. First, we did not measure in advance the participants' ability to remember new words. Thus, there might have been some differences in ability to remember new words between groups from the beginning. Second, we also did not measure their proficiency of the target language. It was unclear whether the participants truly had not known the target English words before the experiment. If some participants already knew some of the target words, this would affect the results.

To reinforce the finding of this study, a further study should examine what types of indirect effects of testing learners benefited from in the restudy session. A questionnaire survey is one of the best ways to deal with this issue. Some typical example of the questions are asking what participants thought when they got a lower score in the first test or what they focused on in the second learning session.

6 Conclusions

The main purpose of this study was to examine whether the indirect effects are useful even if learners experience lower score in a test (Initial Test). The answer is affirmative. This study revealed that the S-TS group had a better score than the S-ST group in Posttests 1 and 2, suggesting that learners can benefit from the indirect effects of testing even if they are not successful in the test. Such effects are powerful tools to establish a strong connection between L1 and L2 pairs. English teachers should be recommended to give their students the test before the restudy session. It was found that tests make a crucial contribution to long-term L2 vocabulary retention even if learners experience poor performance in the test, as long as a restudying session soon after a test is given.

The findings of this study are applicable to the actual English classroom. When teachers organize vocabulary quizzes and design their lessons, they should take the indirect effects of testing into consideration. After the test, the teachers should give their students adequate time to understand the test results, and should conduct an immediate restudy session. If this procedure is followed, the students are able to make better use of the indirect effects of testing even if their test scores are low.

References

- Agarwal, P.K., Finley, J.R., Rose, N.S., & Roediger, H.L. (2017). Benefits from retrieval practice are greater for students with lower working memory capacity. *Memory*, 25(6), 764–771. <https://doi.org/10.1080/09658211.2016.1220579>.
- Arnold, K.M., & McDermott, K.B. (2013). Test-potentiated learning: Distinguish between direct and indirect effects of tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(3), 940–945. <https://doi.org/10.1037/a0029199>.
- Barcroft, J. (2007). Effects of opportunities for word retrieval during second language vocabulary learning. *Language Learning*, 57(1), 35–56. <https://doi.org/10.1111/j.1467-9922.2007.00398.x>.

- Blunt, J.R., & Karpicke, J.D. (2014). Learning with retrieval-based concept mapping. *Journal of Education Psychology, 106*(3), 849–858. <https://doi.org/10.1037/a0035934>.
- Brame, C.J., & Biel, R. (2015). Test-enhanced learning: The potential for testing to promote greater learning in undergraduate science courses. *CBE-Life Sciences Education, 14*, 1–12. <https://doi.org/10.1187/cbe.14-11-0208>.
- Brown, P.C., Roediger, H.L., & McDaniel, M.A. (2014). *Make it stick: The science of successful learning*. Cambridge, MA: Harvard University Press.
- Butler, A.C. & Roediger, H.L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology, 19*, 514–527. <https://doi.org/10.1080/09541440701326097>.
- Carpenter, S.K., Pashler, H., & Vul, E. (2006). What types of learning are enhanced by a cued recall test? *Psychonomic Bulletin and Review, 13*(5), 826–830. <https://doi.org/10.3758/BF03194004>.
- Carpenter, S.K., Pashler, H., Wixted, J.T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory and Cognition, 36*(2), 438–448. <https://doi.org/10.3758/MC.36.2.438>.
- Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory and Cognition, 20*, 633–642. <https://doi.org/10.3758/BF03202713>.
- Craik, F.I.M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology, 104*, 268–294. <https://doi.org/10.1037/0096-3445.104.3.268>.
- Cull, W.L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology, 14*, 215–235. [https://doi.org/10.1002/\(SICI\)1099-0720\(200005/06\)14:3%3C215::AID-ACP640%3E3.0.CO;2-1](https://doi.org/10.1002/(SICI)1099-0720(200005/06)14:3%3C215::AID-ACP640%3E3.0.CO;2-1).
- Halamish, V., & Bjork, R.A. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(4), 801–812. <https://doi.org/10.1037/a0023219>.
- Hays, M.J., Kornell, N., & Bjork, R.A. (2013). When and why a failed test potentiates the effectiveness of subsequent study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(1), 290–296. <https://doi.org/10.1037/a0028468>.
- Kanayama, K., & Kasahara, K. (2015). The effect of word retrieval on L2 vocabulary learning. *Journal of the Hokkaido English Language Education Society, 15*, 21–33.
- Karpicke, J.D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General, 138*(4), 469–486. <https://doi.org/10.1037/a0017341>.
- Karpicke, J.D. (2012). Retrieval-based learning: Active retrieval promotes meaningful learning. *Current Directions in Psychological Science, 21*, 157–163. <https://doi.org/10.1177/0963721412443552>.

- Karpicke, J.D., & Grimaldi, P.J. (2012). Retrieval-based learning: A perspective for enhancing meaningful learning. *Educational Psychology Review*, *24*, 401–418. <https://doi.org/10.1007/s10648-012-9202-2>.
- 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. <https://doi.org/10.1016/j.jml.2006.09.004>.
- Kornell, N., Hays, M.J., & Bjork, R.A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 989–998. <https://doi.org/10.1037/a0015729>.
- McDaniel, M.A., Anderson, J.L., Derbish, M.H., & Morrisette, N. (2007a). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, *19*(4/5), 494–513. <https://doi.org/10.1080/09541440701326154>.
- McDaniel, M.A., Roediger, H.L., & McDermott, K.B. (2007b). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin and Review*, *14*(2), 200–206. <https://doi.org/10.3758/BF03194052>.
- Meyer, A.N.D., & Logan, J.M. (2013). Taking the testing effect beyond the college freshman: Benefits for lifelong learning. *Psychology and Aging*, *28*(1), 142–147. <https://doi.org/10.1037/a0030890>.
- Nation, I.S.P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Putnam, A.L., Nestojko, J.F., & Roediger, H.L. (2016). 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). Oxford, UK: Routledge.
- Richland, L.E., Kornell, N., & Kao, L.S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied*, *15*(3), 243–257. <https://doi.org/10.1037/a0016496>.
- Roediger, H.L., Agarwal, P.K., McDaniel, M.A., & McDermott, K.B. (2011a). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied*, *17*, 382–395. <https://doi.org/10.1037/a0026252>.
- Roediger, H.L., & Karpicke, J.D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*, 181–210. <https://doi.org/10.1111/j.1745-6916.2006.00012.x>.
- Roediger, H.L., & Karpicke, J.D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>.
- Roediger, H.L., & Marsh, E.J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 1155–1159. <https://doi.org/10.1037/0278-7393.31.5.1155>.
- Roediger, H.L., Putnam, A.L., & Smith, M.A. (2011b). Ten benefits of testing and their applications to educational practice. In J. Mestre & B. Ross (Eds.), *Psychology of learning and motivation* (Vol. 55, pp. 1–36). Oxford, UK: Elsevier.

- Soderstrom, N.C., & Bjork, R.A. (2014). Testing facilitates the regulation of subsequent study time. *Journal of Memory and Language*, *73*, 99–115. <https://doi.org/10.1016/j.jml.2014.03.003>.
- Son, L.K., & Kornell, N. (2008). Research on the allocation of study time: Key studies from 1980 to the present (and beyond). In J. Dunlosky & R.A. Bjork (Eds.), *A hand-book of memory and metamemory* (pp. 333–351). Hillsdale, NJ: Psychology Press.
- Tullis, J.G., Finley, J.R., & Benjamin, A.S. (2013). Metacognition of the testing effect: Guiding learners to predict the benefits of retrieval. *Memory and Cognition*, *41*, 429–442. <https://doi.org/10.3758/s13421-012-0274-5>.
- Vojdanoska, M., Cranney, J., & Newell, B.R. (2010). The testing effect: The role of feedback and collaboration in a tertiary classroom setting. *Applied Cognitive Psychology*, *24*, 1183–1195. <https://doi.org/10.1002/acp.1630>.