SOCIETY FOR THE TEACHING OF PSYCHOLOGY

Teaching of Psychology
39(3) 190-193
© The Author(s) 2012
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0098628312450432
http://top.sagepub.com
SAGE

The Testing Effect: Illustrating a Fundamental Concept and Changing Study Strategies

Gilles O. Einstein[1], Hillary G. Mullet[1], and Tyler L. Harrison[2]

Abstract

An important recent finding is that testing improves learning and memory. In this article, the authors describe a demonstration that illustrates this principle and helps students incorporate more testing into their learning. The authors asked students to read one text using a Study–Study strategy and one text using a Study–Test strategy. One week later, the authors tested students' memory for both texts with short-answer quizzes. The results revealed the standard testing effect and served as the basis for a laboratory report that required students to analyze and interpret the results and to answer questions about the testing effect and the experimental design. At the end of the term, students indicated that they were engaging in more testing during their studying.

Keywords

testing effect, meta-cognition

In recent years, psychologists have developed an impressive body of results showing that introducing testing into one's learning produces powerful benefits for memory—benefits that exceed those produced by comparable amounts of time engaging in additional study (see Roediger & Karpicke, 2006a). Interestingly, the available evidence indicates that college students are unaware of the mnemonic benefits of self-testing. Karpicke, Butler, and Roediger (2009) found that only 11% of students reported self-testing as a study strategy and only 1% listed it as their top strategy. By contrast, 84% listed rereading as a study strategy and 55% listed it as their top strategy. In this article, we describe a laboratory that illustrates the importance of testing for long-term retention and helps students incorporate this principle into their own learning.

Roediger and Karpicke's (2006b) research is a particularly clear example of the benefits of testing for memory. They presented participants with prose passages to learn and varied whether they (a) studied one passage for 7 min and then studied it again for 7 min (Study–Study condition) or (b) studied it for 7 min and then tested their memory for it for 7 min (Study–Test condition). During the Test phase in the Study–Test condition, participants simply recalled as much as they could and did not get feedback on their recall. Participants then recalled the passages 5 min, 2 days, or 1 week later. The interesting finding is that memory was nominally lower in the Study–Test condition at the 5-min delay but significantly higher at the 2-day and 1-week retention intervals. Thus, even though both groups spent the same amount of time engaged with the material, testing dramatically lessened forgetting. Indeed, forgetting from 5-min recall to 1-week recall was 35% in the Study–Study condition and only about half of that (18%) in the Study–Test condition.

Recent research also shows that testing produces benefits for academic performance on materials and formats that students typically encounter in college courses. McDaniel, Anderson, Derbish, and Morrisette (2007) encouraged students to use a course website to review materials from the textbook and generally found that test performance was higher for quizzed items than for nonquizzed items. Moreover, they found more pronounced benefits on test performance from short-answer quizzing (quizzing that requires more effortful retrieval) than from multiple-choice quizzing.

Current research shows similar benefits from self-testing. In two experiments with college students, McDaniel, Howard, and Einstein (2009) compared the effectiveness of a 3R (read, recite, review) strategy for reading educational texts with rereading and note-taking strategies. They found powerful benefits of self-testing on measures of immediate and delayed memory and on problem-solving tests.

One implication of this research is that professors should introduce more quizzing into their courses. As one example, Roediger now devotes the last 10 min of his undergraduate courses to testing students on that day's reading assignment and lecture (Elmes, 2010). Another implication is that we should do

[1] Department of Psychology, Furman University, Greenville, SC, USA
[2] Department of Psychology, Georgia Institute of Technology, Atlanta, GA, USA

Corresponding Author:
Gilles O. Einstein, Department of Psychology, Furman University, Greenville, SC 29613, USA
Email: gil.einstein@furman.edu

more to teach students that self-testing is an effective strategy and to encourage them to engage in self-testing during their reading and studying. Toward this end, we developed an exercise for an upper-level Memory and Cognition laboratory that was closely based on Roediger and Karpicke's (2006b, Experiment 1) research.

The major goals of this laboratory project were to illustrate the testing effect and encourage students to introduce more testing into their learning. It is unclear why students prefer rereading as a study strategy over self-testing even though rereading has questionable benefits for memory (McDaniel & Callender, 2008). Because to-be-learned material is immediately accessible when rereading, one explanation is that rereading imbues students with an illusion of confidence regarding their learning (Koriat & Bjork, 2005). Along with this, self-testing is associated with high effort and a lack of fluency, and students may not realize that this struggle improves memory. Thus, participating in a demonstration may be especially important for getting students to apply self-testing to their own learning. On the basis of research showing that directly experiencing the benefits of a strategy makes it more likely that it will be used (Bjork, Storm, & deWinstanley, 2010; Murphy, Schmitt, Caruso, & Sanders, 1987), we attempted to develop a classroom exercise that illustrates the testing effect. Another goal was to create actual data to help students practice their data analysis skills. Although developed for a formal laboratory, the essential elements of this project are also appropriate for a class demonstration.

Method

Participants and Design

Participants were students enrolled in two sections of an upper-level psychology course titled Memory and Cognition. There were 26 participants in each of two classes (22 women in the first class and 18 women in the second class), and these classes were taught in consecutive years.

Materials

The study materials were two short prose passages used by Roediger and Karpicke (2006b) and taken from a test-preparation book for the Test of English as a Foreign Language (Rogers, 2001).[1] The passages were printed on separate sheets of paper, and the order was counterbalanced among students such that each passage was used equally often in the two conditions. For each passage, we created a 12-item short-answer quiz.[2]

Procedure

This activity took place during two laboratory sessions that occurred a week apart. We varied study strategy (Study–Study, Study–Test) within subjects, and we counterbalanced the order of performing these strategies across the two classes. Thus, all students in a given class performed the tasks in the same order, and this allowed us to conduct the demonstration in one classroom. Specifically, all participants in the first class performed the Study–Study condition first and the Study–Test condition second. This order was reversed for the second class. For the Study–Study condition, we asked students to read one of the passages for a 4-min period and then to reread it for a second 4-min period. For both of these periods, we told students that they could highlight, underline, or take notes. For the Study–Test condition, we asked students to read the other passage for 4 min, while again feeling free to highlight, underline, or take notes. For the 4-min test period, students turned the appropriate passage facedown and wrote as much as they could remember about this passage, without concern for exact wording or order, on the backside of the page. After completing their processing of both passages, we asked students to rate how well they thought that they had learned the material from the Study–Study passage and from the Study–Test passage on a scale from 1 (*poorly*) to 5 (*very well*).

At the beginning of the second lab session, we administered surprise quizzes on the passages. We randomly determined the order of testing the passages and used the same order for all students. We allowed students 7 min to take the first quiz and then the same amount of time for the second quiz. Next, students scored their own quizzes as we read the correct answers aloud, stopping as a class to resolve any discrepancies. Students tallied their number of correct answers (of the 12) at the top of each passage, and we then explained the rationale for the laboratory and presented a 20-min PowerPoint lecture that reviewed and discussed some of the experiments described in this article.

We made copies for students of the class scores and gave the students a laboratory report assignment to turn in the following week. For this assignment, students determined the design of the experiment, analyzed the short-answer quiz data, graphed the results, and described limitations of our design.[3]

Finally, to evaluate whether this lab activity produced long-term changes in students' study habits, at the end of the semester we asked students to rate how often they incorporated testing in their reading and studying compared to the beginning of the semester, on a scale from 1 (*much less*) to 5 (*much more*). Students turned these in anonymously.

Results

Students' short-answer quiz scores from the second week indicated that the laboratory produced the testing effect. We included the quiz scores in a 2×2 mixed analysis of variance (ANOVA) with the within-subjects variable of condition (Study–Study, Study–Test) and the between-subjects variable of class (Study–Study first, Study–Test first). This analysis revealed a significant main effect of condition, $F(1, 50) = 6.55$, $\text{MSE} = 2.59$, $\eta_p^2 = .12$, such that performance was higher in the Study–Test condition ($M = 7.10$, $SD = 2.15$) than the Study–Study condition ($M = 6.29$, $SD = 2.09$). Importantly, the analysis revealed no significant main effect of class $F(1, 50) = 1.19$, $\text{MSE} = 6.31$, $p = .28$, and no significant interaction between condition and class, $F(1, 50) = 2.51$, MSE $= 2.59$, $p = .12$.

During the first lab period, students rated how well they thought they learned the material from the two passages.[4] We included these ratings in another 2 × 2 mixed ANOVA like the one described above. Interestingly, there was no main effect of condition, indicating that participants did not perceive a difference in how well they learned the material from the Study–Study ($M = 3.82$, $SD = 0.79$) and Study–Test ($M = 3.98$, $SD = 0.88$) passages, $F < 1$. There was also no significant main effect of class, $F < 1$, and no significant interaction between condition and class, $F(1, 49) = 2.92$, MSE $= 0.64$, $p = .09$. Thus, immediately after studying, students seemed unaware of the mnemonic benefits of testing.

Students' ratings at the end of the semester, however, suggested long-term changes in their study habits. For both classes, we performed one-sample t-tests to examine the difference from a rating of 3 (which represented no change in self-testing).[5] Students in the first class rated themselves as now significantly more likely to use testing both when reading, $t(25) = 6.02$, $p < .001$, $d = 1.18$, and when studying, $t(25) = 9.60$, $p < .001$, $d = 1.88$. Similarly, students in the second class also rated themselves as now significantly more likely to use testing when reading $t(22) = 6.26$, $p < .001$, $d = 1.30$, and when studying, $t(22) = 6.67$, $p < .001$, $d = 1.39$. Examining the combined classes on a descriptive level, 67% of the students rated themselves as somewhat more or much more likely to use testing when reading, and 82% rated themselves as somewhat more or much more likely to use testing when studying, compared to at the beginning of the semester.

We did not evaluate students' learning of the concept that self-testing improves memory in the first class, but we did so in the second class. Prior to performing the exercise, 36% of the students correctly answered a multiple-choice question concerning the benefits of self-testing for long-term memory. A significantly greater percentage of the students (92%) answered the question correctly 1 week after the exercise, $t(24) = 5.53$, $p < .001$, $d = 1.10$.

Discussion

This laboratory produced a significant testing effect, thereby demonstrating that Roediger and Karpicke's (2006b) methodology can be successfully adapted for the classroom. Even though we used a short-answer test (instead of Roediger & Karpicke's free-recall test), it is interesting to note that our effect size (medium) was in the same range as theirs, and the effect was sufficiently robust to be detected with two small classes. Although the order of presenting the Study–Study and Study–Test conditions did not affect performance,[6] we recommend presenting the Study–Study condition first because this reduces the possibility that students will engage in self-testing during the Study–Study phase.

Our multiple-choice item clearly indicated that the exercise enhanced students' understanding of the benefits of testing for memory. However, another goal of this laboratory was to encourage students to introduce more testing into their reading and studying. Existing research suggests that students will sometimes engage in testing during their studying but mainly for diagnosing whether or not they know certain material and not as means of improving their learning and memory (Kornell & Sun, 2009). Students seem to be unaware that retrieval itself enhances memory (Karpicke & Blunt, 2011), and this is consistent with our initial student ratings showing that students perceived that they learned the material equally well in the Study–Study and Study–Test conditions. The results demonstrate, however, that participating in the laboratory or the subsequent lecture had large effects on their study strategies such that most of them reported that they were now more likely to test themselves while reading and studying.

Another goal of this project was to provide students with real data that they could use to refresh and develop their data analysis and interpretation skills. An 8-credit research methods and statistics sequence was a prerequisite for this course, and an important objective of the laboratory component of the course was to help students use and transfer their design and statistical knowledge to varied and novel situations.

In summary, this article presents a straightforward adaptation of the Roediger and Karpicke (2006b) experiment for laboratory or classroom settings. This demonstration and associated lecture illustrated a fundamental principle in cognitive psychology and had pronounced effects on students' reported use of self-testing as a study strategy. Although we cannot be certain whether it was the demonstration per se or the combination of the demonstration and the lecture that affected students' study strategies, several students' comments on the lab report indicated that the demonstration was a compelling factor in changing their behavior.

Declaration of Conflicting Interests

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

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Notes

1. The Sun and Sea Otter passages can be found at http://psych.wustl.edu/memory/stimuli/Stimuli-Roediger&Karpicke2006b.pdf.
2. The short-answer questions and answers can be found at http://www2.furman.edu/academics/psychology/FacultyandStaff/Einstein/Pages/TeachingEffectDemo.aspx.
3. The laboratory assignment can be found at http://www2.furman.edu/academics/psychology/FacultyandStaff/Einstein/Pages/TeachingEffectDemo.aspx.
4. One student from the first class did not complete this rating.
5. Three students from the second class did not complete this rating.
6. Roediger and Karpicke (2006b) also found that counterbalancing order did not affect performance on either initial or delayed tests (J. D. Karpicke, personal communication, November 30, 2010).

References

Bjork, E. L., Storm, B. C., & deWinstanley, P. A. (2010). Learning from the consequences of retrieval: Another test effect. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A Festschrift in honor of Robert A. Bjork* (pp. 351–368). New York, NY: Psychology Press.

Elmes, D. G. (2010). Remembering and researching the old and the new: An interview with Roddy Roediger. *Teaching of Psychology*, *37*, 216–222.

Karpicke, J. D., & Blunt, J. R. (2011, January 20). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science Express*, *331*, 772–775. doi: 10.1126/science.1199327

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

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 187–194.

Kornell, N., & Son, L. K. (2009). Learners' choices and beliefs about self-testing. *Memory*, *17*, 493–501.

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, *19*, 494–513.

McDaniel, M. A., & Callendar, A. A. (2008). Cognition, memory, and education. In H. L. Roediger (Ed.), *Cognitive psychology of memory, Vol. 2 of Learning and memory: A comprehensive reference* (pp. 819–843). Oxford, England: Elsevier.

McDaniel, M. A., Howard, D. C., & Einstein, G. O. (2009). The read–recite–review study strategy: Effective and portable. *Psychological Science*, *20*, 516–522.

Murphy, M. D., Schmitt, F. A., Caruso, M. J., & Sanders, R. E. (1987). Metamemory in older adults: The role of monitoring in serial recall. *Psychology and Aging*, *2*, 331–339.

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.

Rogers, B. (2001). *TOEFL CBT success*. Princeton, NJ: Peterson's.