Increasing Retention without Increasing Study Time

Doug Rohrer [1] and Hal Pashler [2]

[1] University of South Florida
[2] University of California, San Diego

ABSTRACT - Because people forget much of what they learn, students could benefit from learning strategies that provide long-lasting knowledge. Yet surprisingly little is known about how long-term retention is most efficiently achieved. Here we examine how retention is affected by two variables: the duration of a study session and the temporal distribution of study time across multiple sessions. Our results suggest that a single session devoted to the study of some material should continue long enough to ensure that mastery is achieved but that immediate further study of the same material is an inefficient use of time. Our data also show that the benefit of distributing a fixed amount of study time across two study sessions – the spacing effect – depends jointly on the interval between study sessions and the interval between study and test. We discuss the practical implications of both findings, especially in regard to mathematics learning.

Address correspondence to Doug Rohrer, Psychology, PCD4118G, University of South Florida, Tampa, FL 33620, USA. E-mail: drohrer@cas.usf.edu.

Acknowledgements. Our collaborators include Shana Carpenter, Nicholas Cepeda, Noriko Coburn, Michael Mozer, Kelli Taylor, Ed Vol, and John Wixted. This research was supported by the Institute of Education Sciences, U.S. Department of Education (Grant #R305H040108).

Although most people have spent thousands of hours in the classroom, the result of this effort is often surprisingly disappointing. Indeed, both the popular press and the academic literature are replete with examples of educational failure among students and recent graduates. In one assessment of U.S. eighth graders, only 50% were able to correctly multiply -5 and -7 (Reese, Miller, Mazzeo, & Dossey, 1997), and a recent survey of young adults in the U.S. revealed that most could not select the continent in which Sudan is located (National Geographic, 2006). While such findings are partly explained by the fact that some students never learned the information in the first place, we believe that forgetting is often the cause.

For this reason, it seems important to define learning strategies that can promote long-lasting retention. Yet surprisingly little is known about the long-term effectiveness of most learning strategies. For this reason, we have been conducting learning experiments in which subjects are tested as much as one year after the final study session. In a further nod to ecological validity, our subjects learn the kinds of material that people often try to learn, such as vocabulary, geography, foreign language, and mathematics (e.g., Pashler, Rohrer, Cepeda, & Carpenter, in press). In this review, we focus on two decisions that all learners face: how long should one study the same material before quitting or shifting to different material, and how should a fixed amount of study time be distributed across study sessions?

OVERLEARNING

When learners choose to devote an uninterrupted period of time to learning some material or a skill, they must decide when to quit, regardless of whether they later return to the same material. For example, once a student has cycled through a list of vocabulary words until each definition has been correctly recalled exactly one time, the student must decide whether to cycle again through the same list. The continuation of study *immediately* after the student has achieved error-free performance is known as *overlearning*. Many educators argue that overlearning is an effective way to boost long-term retention, and overlearning appears to be quite common in schools. In mathematics courses, for instance, assignments typically include many problems of the same kind, thereby ensuring that students devote much of their study time to overlearning.

Does Overlearning Produce Long-Lasting Benefits?

At first glance, the heavy reliance on overlearning might be seen as consistent with the results of nearly 80 years of empirical literature. In these experiments, subjects either quit or continued studying after some criterion was reached, and the additional study typically boosted subsequent test performance (see Driskell, Willis, & Cooper, 1992, for a meta-analysis). Yet a closer examination of the literature led us to wonder whether the benefits of overlearning might be short-lived. In most overlearning studies, the test was given within a week of the study session, and, in many cases, within an hour. To determine how the benefits of overlearning hold up over meaningful periods of time, we have been measuring the effects of overlearning after various *retention intervals* (RI), the interval between study and test. For example, in one of our experiments (Rohrer, Taylor, Pashler, Wixted, & Cepeda, 2005), subjects learned vocabulary by cycling through a list of word-definition pairs (e.g., *cicatrix-scar*) by repeatedly testing themselves (*cicatrix* - ?, …, *scar*), as one would do with flashcards. They completed either 5 learning trials (Adequate Learning) or 10 learning trials (Overlearning). Adequate Learners generally had no more than one perfect study trial, whereas most Overlearners achieved at least three perfect trials. Subjects were tested either one or four weeks later. As shown in Figure 1, overlearning provided noticeable gains at one week, but these gains were almost undetectable after four weeks. Other studies of ours have confirmed this pattern of declining overlearning benefits, although the length of time over which gains remain detectable varies with the details of the procedure (e.g., Rohrer et al., 2005; Rohrer & Taylor, 2006). In summary, then, we see that while overlearning often increases performance for a short while, the benefit diminishes sharply over time.

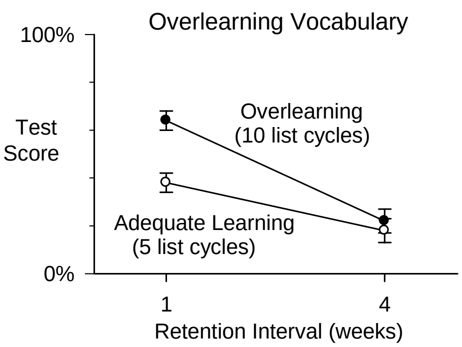

Fig 1. Overlearning. Students learned ten word-definition pairs (e.g., cicatrix-scar) by cycling through the list 5 or 10 times via testing with feedback (cicatrix-?, …, scar). On the subsequent test, the benefit for the 10 trial condition was large after one week but undetectable after four weeks. Error bars reflect plus or minus one standard error.

Implications

In thinking through the practical implications of our overlearning results, it probably makes sense to focus on the *relative efficiency* of overlearning versus alternative strategies. Because overlearning requires more study time than *not* overlearning, the critical question is how the benefits of overlearning compare to the benefits resulting from some alternative use of the same time period. As we will see in the second part of this paper, it seems very likely that devoting this study time to the review of materials studied weeks, months, or even years earlier will typically pay far greater dividends than the continued study of material learned just a moment ago. In essence, overlearning simply provides very little bang for the buck, as each additional unit of uninterrupted study time provides an ever smaller return on the investment of study time. (We hope it is clear that in questioning the utility of overlearning, we are not suggesting that students reduce their study time, nor are we disparaging the use of drill and practice. Rather, we question the wisdom of providing continued practice on material right after error-free performance has been achieved.)

There are, however, situations in which overlearning is desirable. For instance, overlearning appears to be effective in the short term and therefore might be a fine choice for learners who do not seek long-term retention. In addition, there are situations in which an error or even a delayed response might have dire consequences – say, emergency routines performed by pilots, soldiers, or nurses – and here, overlearning is probably advisable and perhaps even necessary.

SPACING OF LEARNING

Overlearning speaks to one aspect of the broader question of how distribution of study time affects learning. This area has been the focus of research for more than a century (see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006, for a recent review). In most research on this topic, a fixed amount of study time is divided across two sessions that are separated by an *inter-session interval* (ISI). If the ISI equals zero, study time is said to be *massed*. Importantly, the retention interval is always measured from the *second* study session. When tested later, performance is usually much better if the study time is spaced rather than massed – a finding known as the *spacing effect* (e.g., Bahrick, 1979; Bjork, 1979). There are numerous theoretical explanations for the spacing effect, but these are beyond the scope of this article (see Dempster, 1989, for a review).

While the superiority of spacing over massing is well established, less is known about how far apart the study sessions should be spaced to promote long-term retention. For instance, does the duration of the inter-session interval affect memory, and, if so, how? We have begun to seek answers to these questions with experiments using long retention intervals.

Varying the Inter-Session Interval

In our first set of spacing experiments, we varied the Inter-Session Interval separating the two study sessions, and the retention interval was fixed (Cepeda, Mozer, Coburn, Rohrer, Wixted, & Pashler, 2007). In the first of these studies, students studied Swahili-English word pairs. The ISI ranged from 5 minutes to 14 days, and the RI was 10 days. ISI had a very large effect on final-test recall, with the 1-day ISI yielding the best recall (Figure 2). In a second experiment in which subjects learned the names of some obscure objects, we used a *six-month* RI, and varied ISI from 5 minutes to 6 months. Effects were even bigger than in the first study, but the optimal ISI was roughly one month (Figure 2).

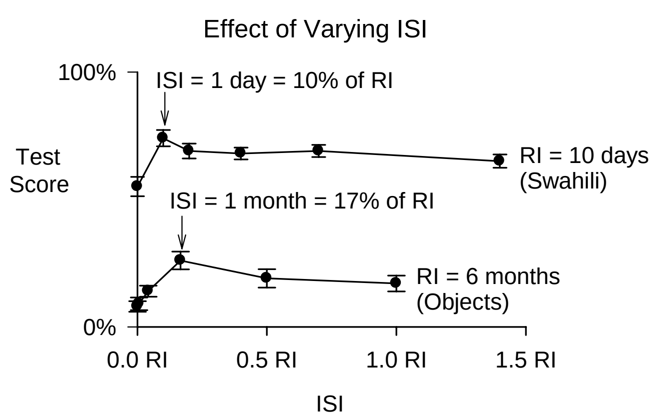

Fig 2. Effect of Varying Inter-Session Interval. In the Swahili experiment, two study sessions were separated by an ISI of 0, 1, 2, 4, 7, or 14 days, followed by a 10-day RI. In the Object Naming experiment, an ISI of 0, 1, 7, 28, 84, or 168 days was followed by a 6-month RI. In both studies, the optimal ISI was about 10-20% of the RI. Error bars reflect plus or minus one standard error.

The Interaction of the ISI and the RI

In comparing the results of the two experiments just described (Figure 2), one sees that the increase in RI from 10 days to six months resulted in an increase in the optimal ISI from about one day to about one month. The results are consistent with an idea that has long been suspected based on studies with short time intervals (Crowder, 1976): that the optimal ISI varies with the RI. To assess this possibility within a single experiment, we are currently conducting a web-based experiment in which we simultaneously vary both ISI (up to 15 weeks) and RI (as long as 50 weeks). Preliminary results from about 1300 subjects indicate that the optimal ISI is indeed varying as expected with RI, with the optimal ISI lying at a value of roughly 10 - 30% of the RI.

The character of this rather intriguing interaction between ISI and RI is illustrated by the hypothetical surface in Figure 3. Here, the vertical axis shows the final test score, with the other two axes representing ISI and RI. Three features are noteworthy. First, for any value of ISI, an increase in RI brings descending performance--the expected forgetting curve. Second, for any value of RI, an increase in ISI causes test score to first increase and then decrease (like the non-monotonic functions in Figure 2). Third, as RI is increased, optimal ISI increases as well, generating a “mountain ridge” that moves gradually outward from the RI axis.

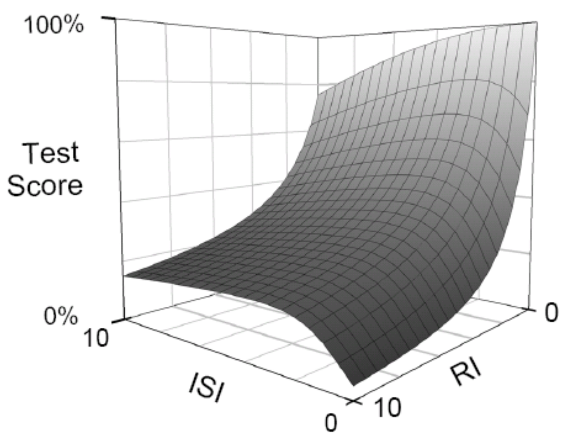

Fig 3. Hypothetical Interaction between ISI and RI. Final test score is shown as a function of Inter-Session Interval and Retention Interval. For any value of ISI, an increase in RI causes test scores to decline monotonically. For any value of RI, an increase in ISI causes test score to first increase and then decrease. The optimal ISI values, which lie along the mountain ridge of the surface, increase as RI increases, producing a mountain ridge that moves gradually outward from the RI axis.

Implications

Our experiments demonstrate that powerful spacing effects occur over practically meaningful time periods. Furthermore, final test performance depends heavily on the duration of the spacing gap, with too-brief gaps causing poorer performance than excessively long gaps. Moreover, spacing effects generally seem to get bigger, not smaller, when one examines longer-term retention. The results have widespread implications for instruction at many levels, of which we will offer just a few examples. Many elementary and middle school teachers present a different set of spelling or vocabulary words each week, but their students might be far better served if material was distributed sporadically across many months. At the college level, instructors often fail to give cumulative final exams, which are likely to induce re-study of material. In the realm of life-long learning, immersion-style foreign language courses are popular, yet their brevity, which prevents sufficient spacing, should produce deceptively high initial levels of learning, followed by rapid forgetting.

MATHEMATICS LEARNING

Because the experiments described thus far required subjects to learn concrete facts, it is natural to wonder whether the results of these studies will generalize to tasks requiring more abstract kinds of learning. To begin to explore this question, we have been assessing the effects of overlearning and spacing in mathematics learning. For example, in one experiment (Rohrer & Taylor, 2006), students were taught a permutation task and then assigned either three or nine practice problems. The additional six problems, which ensured heavy overlearning, had no detectable effect on test scores after one or four weeks. In another experiment with the same task (Rohrer & Taylor, in press), a group of Spacers divided four practice problems across two sessions separated by one week, whereas a group of Massers worked the same four problems in one session. When tested one week later, the Spacers outscored the Massers (74% vs. 49%). Furthermore, the Massers did not reliably outscore a group of so-called Light Massers who worked only half as many problems as the Massers (49% vs. 46%).

This apparent ineffectiveness of overlearning and massing is troubling because these two strategies are fostered by most mathematics textbooks. In these texts, each set of practice problems consists almost entirely of problems relating solely to the immediately preceding material. The concentration of all similar problems into the same practice set constitutes massing, and the sheer number of similar problems within each practice set guarantees overlearning. Alternatively, mathematics textbooks could easily adopt a format that engenders spacing. With this *shuffled format*, practice problems relating to a given lesson would be distributed throughout the remainder of the textbook. For example, a lesson on parabolas would be followed by a practice set with the usual number of problems, but only a few of these problems would relate to parabolas. Other parabola problems would be distributed throughout the remaining practice sets.

The shuffled format not only provides a spaced temporal distribution but also confronts the learner with a variety of problem types within each set, which may itself enhance learning. With the standard format, a lesson on the one-sample *t*-test, for example, is followed by nothing but one-sample *t*-test problems. This provides no discrimination learning to help students determine which features of a problem indicate the appropriate choice of procedure. With a shuffled format, however, problem types are mixed, and students must learn how to find the appropriate strategy for each problem. This benefit seems to be independent of the temporal spacing effect (Rohrer & Taylor, in press).

THE BIGGER PICTURE

Although this brief review has focused on the optimal timing and duration of study, there are, of course, many other decisions learners must make. For example, when preparing for an exam, should students self-test (CASA-?) before seeing the answer (HOUSE), or it is more effective to re-study the answer (CASA-HOUSE)? A sizable body of evidence suggests that *retrieval practice* is usually a wise strategy (e.g., Roediger & Karpicke, 2006), with the caveat that learners receive the correct answer after an error (Pashler, Cepeda, Wixted, & Rohrer, 2005).

Oddly, these kinds of practical questions have mostly been ignored by experimental psychologists over the years (although Harry Bahrick and Robert Bjork are two notable exceptions). Happily, however, there has been a resurgence of interest in this domain in the last few years (see Recommended Readings), and efforts are underway in various places to try to cull the empirical research for simple, concrete principles that can be communicated directly to learners and teachers. Research of this sort should also have spinoffs for educational software. For example, although computer-based instruction typically provides extensive retrieval practice and rapid feedback, it offers a currently unexploited opportunity to schedule study sessions in ways that optimize long-term retention. The various developments currently underway should all help to bring us closer to the time when educational practice will rely chiefly on empirical evidence, rather than on the combination of tradition and fads upon which it has mostly been relying in the past.

Recommended Readings

Bjork, R. A. (1979). Information-processing analysis of college teaching. *Educational Psychologist*, *14*, 15-23.

McDaniel, M.A., Roediger, H.L., & McDermott, K.B. (in press). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review*.

Metcalfe, J., Kornell, N, & Son, L.K. (in press). A cognitive-science based program to enhance study efficacy in a high and low-risk setting. *European Journal of Cognitive Psychology*.

Pashler, H., Rohrer, D., Cepeda, N. J., & Carpenter, S.K. (in press). Enhancing learning and retarding forgetting: Choices and consequences. *Psychonomic Bulletin & Review*.

References

Bahrick, H.P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology: General*, *108*, 296-308.

Bjork, R.A. (1979). Information-processing analysis of college teaching. *Educational Psychologist, 14*, 15-23.

Cepeda, N.J., Mozer, M.C., Coburn, N., Rohrer, D., Wixted, J.T., & Pashler, H. (2007). *Optimizing distributed practice: Theoretical analysis and practical implications*. Unpublished manuscript.

Cepeda, N.J., Pashler, H., Vul, E., Wixted, J.T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*, 354-380.

Crowder, R.G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Dempster, F. N. (1989). Spacing effects and their implications for theory and practice. *Educational Psychology Review, 1*, 309-330.

Driskell, J.E., Willis, R.P., & Copper, C. (1992). Effect of overlearning on retention. *Journal of Applied Psychology, 77*, 615-622.

National Geographic (2006). *Geographic Literacy Study*. Washington, D.C.: National Geographic Society.

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 3-8.

Pashler, H., Rohrer, D., Cepeda, N. J., & Carpenter, S.K. (in press). Enhancing learning and retarding forgetting: Choices and consequences. *Psychonomic Bulletin & Review*.

Reese, C.M., Miller, K.E., Mazzeo, J., & Dossey. J.A. (1997) *NAEP 1996 Mathematics Report Card for the Nation and the States*. Washington, D.C.: National Center for Education Statistics.

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181-210.

Rohrer, D., & Taylor, K. (2006). The effects of overlearning and distributed practice on the retention of mathematics knowledge. *Applied Cognitive Psychology, 20*, 1209-1224.

Rohrer, D., & Taylor, K. (in press). The shuffling of mathematics problems boosts mathematics learning. *Instructional Science.*

Rohrer, D., Taylor, K., Pashler, H., Wixted, J.T, & Cepeda, N.J. (2005). The effect of overlearning on long-term retention. *Applied Cognitive Psychology, 19, 361-374.*