

## *The Effects of Tests on Learning and Forgetting*

SHANA K. CARPENTER, HAROLD PASHLER, JOHN T. WIXTED, and EDWARD VUL  
*University of California, San Diego*

In three experiments, we investigated whether memory tests enhance learning and reduce forgetting more than additional study opportunities. Subjects learned obscure facts (Experiments 1 and 2) or Swahili-English word pairs (Experiment 3) by either completing a test with feedback (Test/Study), or receiving an additional study opportunity (Study). Recall was tested after five minutes, one day, two days, seven days, 14 days, or 42 days. We explored forgetting using an ANOVA, and also by fitting a power function to the data. In all three experiments, testing enhanced overall recall more than restudying. According to the power function, in two out of three experiments testing also reduced forgetting more than restudying, though this was not always the case according to the ANOVA. We discuss the implications of these results for both approaches to measuring forgetting, and for the use of tests in promoting long-term retention. The stimuli used in these experiments may be found at <http://www.psychonomic.org/archive>.

Many psychologists have reported that recalling information on a memory test can strengthen future memory for that information (for a review, see Roediger & Karpicke, 2006b; Wheeler & Roediger, 1992). Tests are often found to strengthen memory even more than extra opportunities to study the material. For example, Cull (2000) had subjects learn obscure English words by pairing them with more common English words that had similar meanings e.g., *handsel: payment*). These word pairs were learned through a test with feedback, or an additional study opportunity. The test with feedback involved an attempt to recall one of the words using the other one as a cue (e.g., *handsel: \_\_\_\_\_*), followed by a presentation of the correct word (*payment*).

The additional study opportunity involved a presentation of both words again (e.g., *handsel: payment*). On a test several days later covering all the word pairs, memory was significantly better for words learned through the test with feedback compared to additional study. The beneficial effect of testing over restudying—i.e., the *testing effect*—has also been observed even when feedback is not provided (Allen, Mahler, & Estes, 1969; Carpenter & DeLosh, 2005, 2006; Kuo & Hirshman, 1996, 1997).

The testing effect seems to be quite robust, having been observed in studies using various paired-associate tasks involving English words (Carpenter, Pashler, & Vul, 2006), English-Yupik word pairs (Carrier & Pashler, 1992), and English-German word pairs (Izawa, Maxwell, Hayden, Matrana, & Izawa-Hayden, 2005). Further afield, the effect has been obtained for face-name associations (Carpenter & DeLosh, 2005; Landauer & Bjork, 1978), general knowledge facts (McDaniel & Fisher, 1991), text passages (Chan, McDermott, & Roediger, 2006; Roediger & Marsh, 2005), and word lists (Carpenter & DeLosh, 2006; Kuo & Hirshman, 1996, 1997). One recent study even extended the testing effect to a map-learning task (Carpenter & Pashler, 2007).

---

Edward Vul is now at the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology. The research reported here was supported by the Institute of Education Sciences, U. S. Department of Education, through Grant R305H040108 to the University of California, San Diego. The opinions expressed are those of the authors and do not represent views of the Institute or the U. S. Department of Education. Parts of this study were presented at the 2005 Annual Meeting of the Psychonomic Society, Toronto, November 10-13, the 2006 Annual Meeting of the American Educational Research Association, San Francisco, April 7-11, and at the 2007 All Hands Meeting of the Temporal Dynamics of Learning Center, La Jolla, January 7-9. We thank Brian Ilagan for his assistance in collecting obscure facts. Correspondence concerning this article should be addressed to Shana K. Carpenter, Department of Psychology, 0109, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0109 ([scarpenter@ucsd.edu](mailto:scarpenter@ucsd.edu))

Given the potential of tests to enhance learning, a number of psychologists have argued that tests should frequently be used in educational contexts to promote learning, not merely—as is the standard practice—to assess it (Bjork, 1988; Dempster, 1989, 1996; Glover, 1989; McDaniel & Fisher, 1991; Pashler, Rohrer, Cepeda, & Carpenter, 2007). This suggestion is supported by a recent study that successfully extended the testing effect outside of the laboratory and into an online college course. McDaniel, Anderson, Derbish and Morrisette (in press) found that final exam performance of students enrolled in an online Brain and Behavior course benefited significantly when students reviewed the information taught through quizzes as opposed to additional reading.

It seems reasonably clear that testing produces an advantage over simply restudying the material. What is less clear is whether this advantage should be interpreted as an increase in the amount of information initially encoded, or a decrease in the rate at which that information is forgotten over time. Or, perhaps it could be both. Our study was designed to explore this issue.

#### *Past Research on the Effects of Tests on Learning and Forgetting*

Some have explored this issue by comparing the effects of testing to restudying upon two different tests: an immediate test, and a test that is delayed by an interval of up to one week. In four studies, restudying was as beneficial, or *more* beneficial, than testing when retention was measured after several minutes, but testing was more beneficial than restudying when retention was measured after two days (Thompson, Wenger, & Bartling, 1978; Wenger, Thompson, & Bartling, 1980), or seven days (Roediger & Karpicke, 2006a; Wheeler, Ewers, & Buonanno, 2003). From these studies, it has been proposed that testing reduces the rate of forgetting over a matter of days, but does not necessarily increase the

original degree of learning (e.g., see Wheeler et al.).

Other studies have looked at cued recall tasks, with final tests delayed by days (Raffel, 1934; Runquist, 1986a, 1986b, 1987) or weeks (Runquist, 1983; Spitzer, 1939). However, these studies did not compare retention for items in a test condition to those in a restudy condition. Instead, they compared retention for items in a test condition without feedback, to other items in a no-test control condition. As such, it is possible that any apparent differences in the rate of forgetting reflect differences in the amount of overall exposure time rather than testing. Furthermore, in some studies (Runquist, 1983, 1986b), final test accuracy in the no-test control condition was assessed as a proportion of all items in that condition, whereas final test accuracy in the test condition was conditionalized on items having been successfully recalled on the intervening test. Thus, the rate of forgetting in the no-test condition was assessed for all items, whereas the rate of forgetting in the test condition was assessed for just the easier and/or initially better-learned items. Therefore, even if tests do not retard forgetting, items in the test condition could have showed slower forgetting by this definition simply because they were the easier items to begin with.

There is also one study known to the present authors that concludes that tests do not slow down the rate of forgetting. Slamecka and Katsaiti (1988) compared the effects of testing versus restudying upon final tests that were given immediately, after a one-day delay, or after a five-day delay. Across the one-day and five-day delayed tests, there was no significant advantage for testing over restudying, and no interaction between test condition and retention interval. These results are difficult to interpret, however, because contrary to past research on the testing effect, Slamecka and Katsaiti did not observe any significant overall benefit of testing over

restudying. Had the usual learning advantage of testing over restudying been observed, would a significant reduction in the rate of forgetting also have been observed?

In summary, the issue of whether or not testing reduces the rate of forgetting more than restudying appears to require further exploration. Whereas some studies reported that tests appear to slow down forgetting up to seven days (Roediger & Karpicke, 2006a; Wheeler et al., 2003), at least one study that measured retention after several days found that testing did not slow down forgetting (Slamecka & Katsaiti, 1988). Runquist (1983) used a much longer retention interval of 21 days and reported slowing of forgetting by testing, however this result could well have been driven by comparisons between testing versus no re-exposure, which is quite different than the comparison of testing versus restudying. The current study set out to resolve what might be considered equivocal evidence about the role of testing on forgetting up to several days, and to provide much-needed data on the effects of testing vs. restudying on forgetting over even longer time periods.

### *Current Study*

The current study explored the effects of tests on learning and forgetting using a much longer range of time than has ever been explored in past research. Whereas past studies measured retention for tests versus restudy opportunities at two or three points in time spanning a range of up to one week, our study explored memory at six different points in time spanning a range of six weeks. We compared a group of items that was learned through an additional study opportunity to a group of items that was learned through a cued recall test, and then we measured recall for a different subset of items from each condition after five minutes, one day, two days, seven days, 14 days, and 42 days. We used tests of cued recall in order to insure that

individual items could be equated in the amount of study time they received. In Experiments 1 and 2, the items were obscure facts (e.g., *Greyhounds have the best eyesight of any dog*), and in Experiment 3, Swahili-English word pairs (e.g., *Somo: Friend*).

We also controlled for the possibility that items learned through a restudy opportunity could have an "unfair" advantage over items learned through a test. This possibility exists because items in a test condition are not always retrieved with 100% accuracy, whereas in a restudy condition, 100% of the items are presented again. Many past studies have shown that if feedback is not provided, any items that are not recalled on the test have virtually no chance of being learned (Allen et al., 1969, Bjork, 1988; Kuo & Hirshman, 1996, 1997; Modigliani, 1976; Postman & Phillips, 1961), and on cued recall tasks, items eliciting an error not followed by feedback have almost no chance of being recalled after a further delay (Pashler, Cepeda, Wixted, & Rohrer, 2005). Therefore, to avoid handicapping the learning of items in the test condition, we provided feedback after every item on the cued recall test.

In the test condition (referred to as *Test/Study*), subjects were given four seconds to recall the answer to each fact in Experiments 1 and 2 (e.g., What breed of dog has the best eyesight?), at which time they were shown the correct answer (Greyhound) for two additional seconds. For the restudy condition (referred to as *Study*), subjects were presented with both the question and answer (e.g., What breed of dog has the best eyesight? Greyhound) for six seconds. In the Test/Study condition of Experiment 3, subjects were given two seconds to recall the English translation of the Swahili word (e.g., Somo: \_\_\_\_\_), at which time they were shown the complete word pair (Somo: Friend) for two additional seconds. In the Study condition of Experiment 3, subjects were shown a four-second presentation of the complete word pair

(Somo: Friend). Thus, our design, which follows Carrier and Pashler (1992), equates the amount of time subjects spend on each trial for both Test/Study and Study. Furthermore, it is worth noting that the Test/Study procedure, relative to the Study procedure, actually involves a reduced amount of time during which the correct answer is presented. Any benefits conferred by testing, therefore, cannot be sufficiently explained as a function of mere exposure time.

In Experiment 1, we gave subjects one Test/Study or one Study on each fact. Past research shows that the effects of tests are stronger when the number of tests versus restudy opportunities is increased (Allen et al., 1969; Kuo & Hirshman, 1996). Therefore, to magnify any effects produced by tests, Experiment 2 provided subjects with three Test/Study or three Study on each fact. Experiment 3 also provided subjects with three Test/Study or three Study on each Swahili-English word pair.

#### *Analyzing the Rate of Forgetting*

All of the past studies exploring the effects of tests on forgetting have used an Analysis of Variance (ANOVA) to compare forgetting rates. Here, the question was simply whether there was an interaction between test condition and retention interval. According to this approach, the rate of forgetting in Condition A is said to differ from that of Condition B if the difference in percent correct between the two conditions grows larger as the retention interval increases.

The alternative approach to the ANOVA is to compare forgetting rates using a mathematical characterization of the rate of forgetting. There has been a substantial research tradition on this topic, beginning with Ebbinghaus (1885/1913), who first showed that the decline in percent correct per unit of time is greatest at first, and then gradually slows down (see Wixted, 1990; Wixted & Ebbesen, 1991, 1997). One of the best-known

functions to describe the forgetting process is a power function, originally proposed by Wickelgren (1974):

$$y = a(bt + 1)^{-c}$$

Here,  $t$  represents time,  $a$  is a constant representing the degree of original learning (i.e., the proportion of items recalled at  $t = 0$ ),  $c$  is a constant representing the rate of forgetting, and  $b$  is a scaling constant.<sup>1</sup> This power function has been shown to accurately describe a wide range of individual and group data across different memory tasks, and even different species of subjects (e.g., Wickelgren, 1974; Wixted 2004).

The power function analyzes forgetting using a different approach from that of the ANOVA. Here, the power function is fit to the data for each subject, consequently yielding two separate estimates of the forgetting rate parameter ( $c$ ), one for Test/Study and one for Study. These values can then be directly compared. The fit also provides two estimates of the degree of learning parameter ( $a$ ) per subject, one for Test/Study and one for Study, and these two values can be compared to determine whether there is a significant difference in the degree of original learning between the two conditions.

Our study explored differences in the degree of learning and rate of forgetting for Test/Study vs. Study using both the ANOVA-based approach and the power function approach. As will be made clear later, the conclusions from these two approaches do not always seem to coincide. In the General Discussion, we provide more information on why this is the case, and we discuss the implications of each approach in measuring the time course of forgetting.

Our study addressed the following questions: After sampling recall across a six-week time interval and controlling for potential differences in item selection, are tests with feedback more likely than restudy opportunities to (1) increase the degree of

learning, and (2) reduce the rate of forgetting? (3) Are the effects of tests strengthened by providing three Test/Study versus three Study, as opposed to only one Test/Study versus one Study? Finally, (4) Are the effects of tests on learning and forgetting similar across different types of materials—namely, obscure facts and Swahili-English word pairs?

### Experiments 1 and 2

#### *Subjects*

Subjects were drawn from our laboratory's pool of online research subjects. Individuals enrolled in this panel generally access the Internet on a frequent basis, and have shown themselves to be diligent in their previous participation in extended memory experiments. Fifty-five subjects (42 females and 13 males) completed Experiment 1, and 57 different subjects (47 females and 10 males) completed Experiment 2. Subjects in Experiment 1 ranged in age from 19 to 63, with 62% of subjects below the average age of 30.02 ( $SD = 9.75$ ), and 38% above. Subjects in Experiment 2 ranged in age from 18 to 63, with 56% of subjects below the average age of 34.81 ( $SD = 12.19$ ), and 44% above.

Internet testing provides us with a larger, more diverse group than we could have obtained using standard laboratory testing and greatly facilitates the repeated testing required for experiments of this sort. Although laboratory-based experiments are more common in the memory field, we along with other investigators have found consistent patterns of results across laboratory- and web-based experiments (e.g., Birnbaum, 1999; Krantz & Dalal, 2000; McGraw, Tew, & Williams, 2000; Reips, 2002), and it is our impression that our paid Internet panel in particular is generally more careful and attentive than subjects drawn from the typical undergraduate subject pool. Subjects performed six sessions (the first was about 20 to 35 minutes in length, and the remaining five

sessions lasted about two minutes each) in exchange for payment of \$20.00.

#### *Materials and Design*

We used a variety of on-line and printed sources to assemble 60 obscure facts (e.g., *Greyhounds have the best eyesight of any dog; Fake pearls were once made out of fish scales; "Jack" is the most commonly used name in nursery rhymes*). The complete set of stimuli used in all three experiments can be found at <http://www.psychonomic.org/archive>.

We used a 2 x 6 (test condition: Test/Study vs. Study x retention interval: five min, one day, two days, seven days, 14 days, 42 days) within-subjects design. For each subject, five facts were randomly assigned to one of the 12 possible conditions. To measure recall across time, a different group of 10 facts (five from Test/Study and five from Study) was tested after five minutes, one day, two days, seven days, 14 days, or 42 days.

#### *Procedure*

For all data collection, we used a website that ran on the free and open-source LAMP (Linux, Apache, MySQL, and PHP) framework. This website was tested in order to insure its accessibility through all major web-browsers. The experiment was programmed using both server-side PHP scripts and client-side JavaScript. The server-side PHP programs stored data and controlled experiment flow, and the client-side JavaScript precisely controlled the timing of item presentation and recorded response times (see Vul & Pashler, 2007, for timing accuracy details).

Subjects first answered several demographic questions about their gender, age, level of education, and what type of environment they were in while doing the experiment. Subjects indicated their environment by choosing from among several alternatives (e.g., at home in a room by myself, in a library, in an Internet café, etc.).

Subjects then read instructions on the computer screen, telling them that they would be learning obscure facts, and they should try to remember these facts for a later, unspecified memory test. Once subjects began the experiment, each of the 60 facts was presented one at a time, in statement format (e.g., Greyhounds have the best eyesight of any dog) for six seconds. After each fact, a blank screen with a “continue” button appeared, and subjects clicked this button to view the next fact. This procedure was used to increase the chances that subjects would encode each fact without missing the presentation of any item due to any potential distractions. The facts were presented in a different random order for each subject.

Following the presentation phase, all 60 facts were then encountered again. This time, half of them appeared as Test/Study and the other half as Study. For Test/Study, subjects were presented with the fact in question format, and were instructed to covertly recall the correct one-word answer within four seconds. After four seconds, the correct answer was displayed, along with the question, for two additional seconds. Thus, during Test/Study trials, subjects were not required to enter a response, but instead to simply recall the answer in their mind. As such, any benefits conferred by the Test/Study procedure therefore cannot be attributed to the overt response (e.g., typing in an answer, or writing it down), but rather to the act of recalling the response and then receiving feedback. For Study, subjects were presented with the fact in question format, along with the correct answer, for six seconds. Thus, the total time that each item was presented in the Test/Study and Study conditions was always six seconds. In-between the presentation of each item, subjects encountered a blank screen with a “continue” button, which they clicked to view the next item.<sup>2</sup>

For each subject, the order of presentation of all 60 facts, as well as the order in which

each fact appeared as a Test/Study or Study, was randomized. In Experiment 1, each fact was presented once as a Test/Study or a Study. In Experiment 2, each fact was presented three times as a Test/Study or a Study. Each time the facts were repeated in Experiment 2, the same facts were assigned to Test/Study and Study for each subject, and the facts were presented in a new random order for each subject. Upon completing all of the Test/Study and Study trials, subjects in both experiments engaged in a five-minute video game distracter task.

Immediately following the distracter task, subjects were given a final test over 10 of the facts (five from Test/Study and five from Study). For this test, subjects were presented with the fact in question format (What breed of dog has the best eyesight?), and were required to type in the correct one-word answer. They were given unlimited time to respond, and feedback was not provided. Subjects were instructed to guess if they were unsure about the correct answer. Completion of this test marked the end of Session 1. For Sessions 2 through 6, subjects were given the same type of test again, over 10 different facts (five from Test/Study and five from Study).

When the time came for a subject to perform Sessions 2 through 6, a server-side script program sent the subject an e-mail containing a URL linking the subject's computer browser to the server. Session 2 could be completed between 18 and 32 hours following Session 1; Session 3 could be completed between 42 and 56 hours following Session 1; Session 4 between 156 and 192 hours following Session 1; Session 5 between 312 and 384 hours following Session 1; and Session 6 between 984 and 1080 hours following Session 1. In Sessions 2 through 6, a new group of 10 facts each was tested (five from Test/Study and five from Study), and each of these sessions lasted approximately two minutes. The total time to complete all six sessions was approximately 25 minutes in

Experiment 1, and 45 minutes in Experiment 2.

*Results and Discussion*

*Effects of Environment.* The majority of subjects participated while in a room by themselves: at least 67% at each of the six sessions in Experiment 1, and at least 81% in Experiment 2. Differences in environment did not significantly affect final test accuracy, nor did they interact with other variables.

*Learning and Forgetting for Test/Study versus Study.* In both experiments, we analyzed the data using two different methods. In one method, we performed a 2 x 6 (Test Condition x Retention Interval) Repeated Measures ANOVA. According to the ANOVA, the main effect of test condition was significant in Experiment 1 [ $F(1, 54) = 14.36, p < .001, MSE = .032, \eta_p = .21$ ], and in Experiment 2 [ $F(1, 56) = 47.38, p < .001, MSE = .023, \eta_p = .46$ ]. In Experiment 1, facts learned through one Test/Study were retained at an average overall rate of 69%, whereas facts learned through one Study were retained at an average overall rate of 64%. The overall

rate of recall, as well as the advantage of Test/Study over Study, was even greater in Experiment 2. Facts learned through three Test/Study were retained at an average overall rate of 78%, whereas facts learned through three Study were retained at an average overall rate of 70%.

In both experiments, significant forgetting occurred across the six-week time interval. The main effect of retention interval was significant in Experiment 1 [ $F(5, 270) = 168.73, p < .001, MSE = .038, \eta_p = .76$ ], and in Experiment 2 [ $F(5, 280) = 147.51, p < .001, MSE = .036, \eta_p = .72$ ]. The Test Condition x Retention Interval Interaction was not significant in Experiment 1,  $F < 1$ , however, this interaction was significant in Experiment 2 [ $F(5, 280) = 3.88, p < .01, MSE = .023, \eta_p = .06$ ]. The mean proportion of facts recalled in all conditions for Experiment 1 and 2 are reported in the top and middle sections, respectively, of Table 1.

In the other method, we fit the proportion of facts recalled at each of the six intervals to the power function  $y = a(bt + 1)^{-c}$ . The within-subjects manipulation of both test condition

Table 1  
Mean Proportion of Items Retained as a Function of Retention Interval and Test Condition (with standard errors)

Test Condition	Retention Interval						Total
	5 min	1 day	2 days	7 days	14 days	42 days	
Experiment 1 (1 Test/Study vs. 1 Study over obscure facts)							
Test/Study	.93 (.02)	.88 (.02)	.86 (.02)	.66 (.04)	.47 (.04)	.34 (.03)	.69 (.02)
Study	.91 (.02)	.82 (.03)	.78 (.03)	.60 (.04)	.42 (.04)	.30 (.03)	.64 (.02)
Total	.92 (.02)	.85 (.02)	.82 (.02)	.63 (.04)	.44 (.03)	.32 (.02)	
Experiment 2 (3 Test/Study vs. 3 Study over obscure facts)							
Test/Study	.96 (.01)	.93 (.02)	.89 (.02)	.84 (.03)	.68 (.04)	.39 (.03)	.78 (.02)
Study	.93 (.02)	.87 (.03)	.84 (.03)	.70 (.04)	.52 (.04)	.36 (.03)	.70 (.02)
Total	.95 (.02)	.90 (.02)	.86 (.02)	.77 (.03)	.60 (.03)	.38 (.02)	
Experiment 3 (3 Test/Study vs. 3 Study over Swahili-English word pairs)							
Test/Study	.59 (.05)	.43 (.06)	.41 (.06)	.30 (.05)	.25 (.05)	.19 (.05)	.36 (.04)
Study	.51 (.05)	.36 (.05)	.30 (.05)	.21 (.04)	.20 (.05)	.16 (.04)	.29 (.04)
Total	.55 (.05)	.40 (.05)	.35 (.05)	.26 (.04)	.22 (.05)	.18 (.04)	

and retention interval made it possible for us to fit the power function to the data for each individual subject (e.g., see Rickard, 2004, for a discussion of the advantages of individual subject fits over fits to the averaged data). Each subject’s data were fit using Maximum Likelihood Estimation (Myung, 2003) with  $b$ —the scaling constant—constrained to be equal across all subjects and across both test conditions (e.g., see Wixted & Carpenter, 2007). To estimate the value of the scaling constant, we first fit the data that had already been averaged over subjects. The  $b$  value estimate that resulted from this grand average fit was then used across all subjects and all conditions to carry out the individual subject fits. This fitting process was carried out separately for Experiment 1 and Experiment 2. For each experiment, the fit yielded a total of four parameters per subject: (1) degree of learning for Test/Study, (2) degree of learning for Study, (3) rate of forgetting for Test/Study, and (4) rate of forgetting for Study.

Figure 1 shows the average percent of facts recalled across the six retention intervals for Experiment 1 (top panel) and Experiment 2 (bottom panel). The smooth curves represent the average of the 55 individual subject forgetting curves for Experiment 1, and the 57 individual subject forgetting curves for Experiment 2.

We used a binomial sign test to evaluate whether a significant number of subjects exhibited higher degrees of learning for Test/Study than for Study, and lower rates of forgetting for Test/Study than for Study.<sup>3</sup> In Experiment 1, the majority of subjects (60%) showed an advantage in the degree of learning for Test/Study over Study, and 60% also showed a lower rate of forgetting for Test/Study than for Study. These proportions failed to reach significance according to the sign test,  $ps = .18$ . In Experiment 2, however, the sign test indicated that a significant proportion of subjects (72%) showed higher degrees of learning for Test/Study than for

Study,  $p < .01$ , and a significant proportion of subjects (68%) showed lower rates of forgetting for Test/Study than for Study,  $p < .01$ .

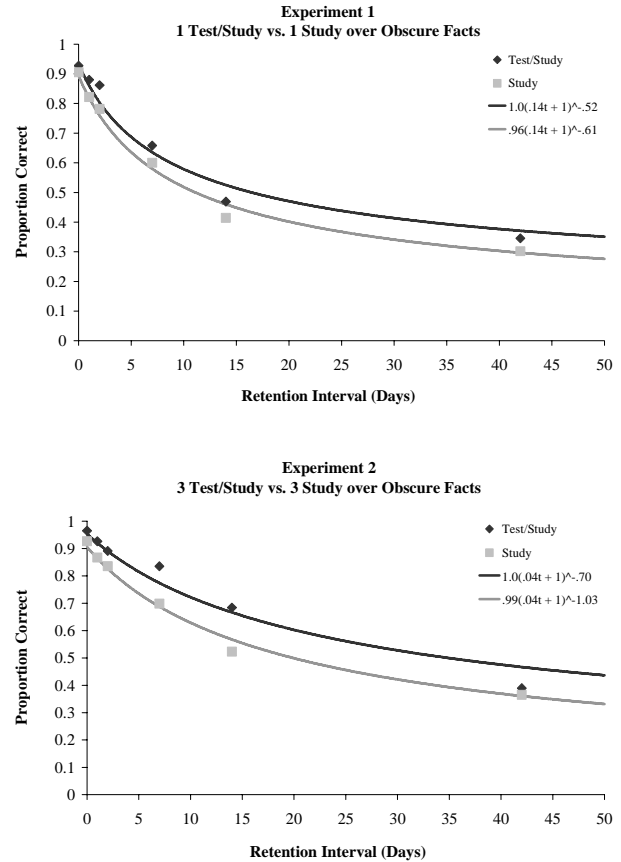


Figure 1. Subjects were given a test with feedback (*Test/Study*) vs. a restudy opportunity (*Study*) over each fact. Recall of these facts was tested after five min, one day, two days, seven days, 14 days, or 42 days following just one *Test/Study* vs. one *Study* (upper panel), or following three *Test/Study* vs. three *Study* (lower panel). The points represent the average proportion of facts recalled from *Test/Study* vs. *Study* at each of the six retention intervals. The power function  $y = a(bt + 1)^c$  was fit to each subject’s data to yield a degree of learning parameter and a rate of forgetting parameter. Having just one *Test/Study* vs. one *Study* increased the degree of learning and reduced the rate of forgetting (upper panel), but these effects did not reach significance. Having three *Test/Study* vs. three *Study* significantly increased the degree of learning, and significantly reduced the rate of forgetting (lower panel). The smooth curves represent the mean of the 55 individual subject forgetting curves in Experiment 1, and the 57 individual subject forgetting curves in Experiment 2. In all three experiments the curve-fitting procedure produced a few extreme parameter estimates for degree of learning and rate of forgetting. These extreme values did not affect the visual display of the graphs, but they did affect the mean parameter estimates. The parameter estimates in the equations, therefore, are medians rather than means.

The apparent difference in the rate of forgetting between Test/Study versus Study may have been influenced by a ceiling effect, however. Close examination of Figure 1 reveals that performance was very high in both conditions at the five-minute retention interval, especially in Experiment 2. If the measure of recall at short retention intervals was not constrained by a relatively easy task, would differences in the rate of forgetting still emerge between Test/Study and Study? Experiment 3 was conducted to explore this issue. Instead of using obscure facts, we used Swahili-English word pairs, which we assumed would be more difficult to recall and less likely to yield a ceiling effect at the five-minute retention interval.

### Experiment 3

#### *Subjects*

Forty-four subjects, drawn from the same pool as in Experiments 1 and 2, completed Experiment 3. None of these subjects had participated in either of the two previous experiments. Subjects performed six sessions (the first was about 15 to 20 minutes in length, and the remaining five sessions lasted about two minutes each), and were paid \$10.00. Subjects (35 females and nine males) ranged in age from 19 to 63, with 59% of subjects below the mean age of 30.57 ( $SD = 10.28$ ), and 41% above.

#### *Materials*

We assembled 60 Swahili-English word pairs (e.g., *Somo-Friend*; *Farasi-Horse*; *Gereza-Jail*). According to Wilson's (1988) norms, the English words were all nouns that ranged between three and seven letters, between one and three syllables, occurred with a frequency of greater than 20 per million, and ranged in concreteness from 400 to 700. The Swahili equivalents for each English word were obtained from Yale University's Kamusi Project website (2005).

#### *Design and Procedure*

All aspects of the design were identical to Experiment 2 except for the materials and the presentation duration. In Experiment 3, subjects were presented with each Swahili-English word pair for a total of four seconds. In the Test/Study condition, subjects saw only the Swahili word (Somo: \_\_\_\_\_) for two seconds, and were instructed to try and covertly recall the correct English word (Friend). After two seconds had elapsed, the complete word pair was presented (Somo: Friend) for two additional seconds. In the Study condition, subjects saw the complete word pair for a total of four seconds.

#### *Results and Discussion*

*Effects of Environment.* At each of the six sessions in Experiment 3, at least 66% of subjects participated while in a room by themselves. Differences in environment did not systematically affect final test accuracy.<sup>4</sup>

*Learning and Forgetting for Test/Study vs. Study.* As in the previous experiments, we analyzed the data for Experiment 3 by using the ANOVA-based method and the curve-fitting method. According to the ANOVA, the main effect of test condition was significant [ $F(1, 43) = 30.22, p < .001, MSE = .023, \eta_p = .41$ ], as was the main effect of retention interval [ $F(5, 215) = 28.41, p < .001, MSE = .057, \eta_p = .40$ ]. As in the previous experiments, items learned through Test/Study (36%) were retained significantly better than items learned through Study (29%), and significant forgetting occurred across the six-week time interval. However, the Test Condition x Retention Interval Interaction was not significant [ $F < 1$ ]. The mean proportions of items recalled in all conditions for Experiment 3 are reported in the bottom section of Table 1.

Figure 2 shows the average percent of words recalled across the six retention intervals for Experiment 3. The smooth curves

represent the average of the 44 individual subject forgetting curves.

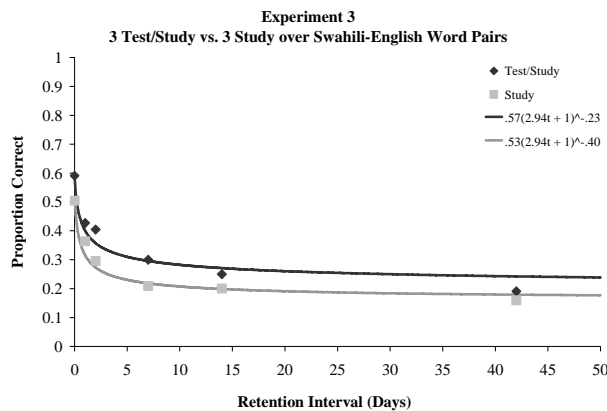


Figure 2. Experiment 3 replicated Experiment 2 using Swahili-English word pairs instead of obscure facts. Having three Test/Study vs. three Study over Swahili-English word pairs enhanced the degree of learning, though this effect was not significant. Having three Test/Study vs. three Study over Swahili-English word pairs significantly reduced the rate of forgetting. The smooth curves represent the mean of the 44 individual subject forgetting curves.

Out of the 44 total subjects, 27 exhibited a higher degree of learning for Test/Study than for Study, 15 exhibited a higher degree of learning for Study, and two exhibited the same degree of learning for Test/Study as for Study. The proportion of subjects showing higher degrees of learning for Test/Study (64%) approached significance,  $p = .09$ . Out of the 44 subjects, 29 exhibited a lower rate of forgetting for Test/Study than for Study, 13 exhibited a lower rate of forgetting for Study, and two exhibited the same rate of forgetting for Test/Study as for Study. The proportion of subjects showing lower rates of forgetting for Test/Study than for Study (69%) was significant,  $p < .05$ .

### General Discussion

In three experiments, Test/Study was more beneficial to memory recall than Study. We observed a significant benefit on the recall of obscure facts by one Test/Study versus one Study (Experiment 1), and an apparently larger benefit by three Test/Study versus three

Study (Experiment 2). Furthermore, we observed the same benefit when comparing three Test/Study versus three Study on the recall of more difficult Swahili-English word pairs (Experiment 3). These results are consistent with a number of prior studies showing that testing enhances memory as assessed through cued recall more than restudying (e.g., Carpenter & DeLosh, 2005, 2006; Carpenter et al., 2006; Carrier & Pashler, 1992; Izawa, 1992).

This finding was observed despite the fact that we equated the total study time in the Test/Study and Study conditions. Indeed, the advantage of the Test/Study procedure was actually produced by depriving the subject of certain information (i.e., the one-word answer to the fact query, or the correct English word for the Swahili cue) for two seconds, so the amount of study time that the subjects had available with all of the information "in hand" was actually greater in the Study condition than in the Test/Study condition. The results also suggest that the benefit of Test/Study over Study is amplified by increasing the number of tests versus restudy opportunities, as some earlier investigators have observed (Allen et al., 1969; Kuo & Hirshman, 1996).

What do our results say about the effects of tests on the rate of forgetting? According to the curve-fitting analysis, Experiments 2 and 3 revealed a significant reduction in the rate of forgetting due to testing. This result is in line with previous studies that have reported reductions in forgetting due to free recall testing (e.g., Roediger & Karpicke, 2006a; Thompson et al., 1978; Wenger et al., 1980; Wheeler et al., 2003). According to the ANOVA-based analysis, however, only Experiment 2 revealed a significant reduction in the rate of forgetting due to testing. According to the ANOVA-based approach, therefore, our results from Experiments 1 and 3 might agree with those of Slamecka and Katsaiti (1988), who failed to observe the Test x Retention Interval Interaction and therefore

concluded that tests do not slow down the rate of forgetting. The conclusions drawn from the current data depend on which approach one takes to measuring the rate of forgetting.

### *Different Approaches to Measuring Forgetting*

In this paper we have presented two different approaches to measuring the rate of forgetting. It is interesting that the results of the ANOVA-based approach do not always appear to agree with those of the curve-fitting approach. Specifically, in Experiments 1 and 3, the Test x Retention Interval Interaction was not significant, suggesting that the rate of forgetting for Test/Study did not differ from that of Study. According to the curve-fitting approach, however, the rate of forgetting was consistently lower for Test/Study than for Study. Although this difference did not reach significance in Experiment 1 according to a binomial sign test, the difference was significant in Experiments 2 and 3.

How can it be that tests slow down the rate of forgetting according to the curve-fitting analysis, but this is not always the case according to the ANOVA-based analysis? A non-significant interaction, according to the ANOVA-based method, indicates that the numerical advantage of Test/Study over Study remains relatively constant across the six retention intervals. This relative constancy is reflected in one aspect of the curve-fitting analysis in Experiment 3, namely the fact that the forgetting curves for Test/Study vs. Study seem to differ by only a constant amount on the y-axis. At first glance, “parallel” curves like these may lead one to assume that the two forgetting rates do not differ.

However, “parallel” curves in this case would imply no difference in the rate of forgetting only if one assumes that forgetting curves project towards an asymptote greater than zero. That is, even if the lower curve (Study) projects towards an asymptote of zero, the upper curve (Test/Study) would have to

project towards an above-zero asymptote in order for the relatively constant advantage of Test/Study over Study to remain constant beyond the range of intervals studied. Therefore, according to the ANOVA-based method, a non-significant interaction is evidence that the two forgetting functions do not differ provided that—in at least one of the conditions—forgetting functions decline to an asymptote greater than zero.

The curve-fitting method, on the other hand, assumes that memory performance in both conditions is projecting towards an asymptote of zero. If this is true (e.g., see Anderson, 2000; Wickelgren, 1974; Wixted, 2004; Wixted & Carpenter, 2007), then observing a relatively equal sized advantage of Test/Study over Study at each retention interval actually favors the conclusion that the forgetting rates differ. If there were no difference in the rate of forgetting, the two forgetting curves would begin at different points on the y-axis and then rapidly converge together (e.g., see the simulations by Loftus, 1985, 2002). Thus, in Experiment 3, according to the curve-fitting method, the two nearly “parallel” curves suggest different rates of forgetting. That is, the top curve (Test/Study) will take longer to reach a value arbitrarily close to zero than will the bottom curve (Study). Consider the two forgetting curves in Figure 2. As each curve projects towards zero, the Test/Study curve takes longer than Study to fall below any given level of performance. To fall below 25%, for example, the curve for Study requires about three days, whereas the curve for Test/Study requires about 30 days.

It is possible, therefore, to reconcile the apparent differences in the results of the ANOVA-based approach and the curve-fitting approach by understanding the assumptions that each approach adopts concerning the asymptotic value of forgetting rates. Asymptote is therefore an important point that should be considered in any work that investigates the time course of forgetting.

*Using the ANOVA vs. Power Function to Measure Forgetting*

*Advantages of the ANOVA.* An obvious benefit of using the ANOVA is that it involves concrete and straightforward statistical analyses. The curve-fitting method, on the other hand, requires more complex mathematical analyses using data that are—under ideal circumstances—collected from the same subjects over multiple retention intervals. Collecting data from the same subjects across multiple retention intervals and multiple conditions will undoubtedly result in a limited number of trials per condition (i.e., in the current study, only five items in each of the 12 within-subjects conditions), which can increase the chances that one item will have a non-trivial effect on the mean. Such item effects would seem to reduce the likelihood of detecting significant differences between conditions. The ANOVA is less susceptible to this problem, requiring only two retention intervals to detect an interaction.

Another limitation of the curve-fitting approach is that it does not appear to fit the data well for every individual subject, sometimes resulting in extreme parameter estimates. Presence of these outliers could reflect the existence of a non-normal population of forgetting rate parameters. One implication of using the curve-fitting approach, therefore, is that one may need to evaluate differences between parameter estimates using non-parametric analyses (e.g., the binomial sign test) that do not rely on the assumption that the population of parameter estimates is normally-distributed.

*Advantages of the Power Function.* The power function measures forgetting as a proportional loss of the amount of information that was originally learned, whereas the ANOVA measures the absolute loss of information from memory over time. For example, suppose subjects in Condition A initially recall 22% of the items on a word list,

whereas subjects in Condition B recall 90%. Suppose that after one week subjects in Condition A are able to recall 2% of the words, whereas subjects in Condition B can recall 70%. The difference in the proportion of items retained is 20% in both cases, so the ANOVA would suggest equal rates of forgetting in the two conditions (such a conclusion carries with it the theoretical claim that the two functions project towards different asymptotes). However, an absolute loss of 20% is a far smaller portion of the material that was originally learned in Condition B compared to Condition A. As such, the power function would interpret Condition A as having a much higher rate of forgetting (a conclusion that rests on the assumption that both forgetting functions project towards an asymptote of zero).

The power function well characterizes the curvilinear form of forgetting whereby the amount of information lost from memory is greatest at first, and then gradually slows down with the passage of time. This trend was first described by Ebbinghaus (1885/1913), and has since been supported by several decades of research on forgetting (e.g., Rubin & Wenzel, 1996).

Finally, in measuring the time course of forgetting, it can be useful to know not only whether a difference exists in the rate of forgetting between two conditions, but also exactly what that difference is. As we have discussed, both the ANOVA and power function can detect differences in the rate of forgetting between two conditions. Only the power function, however, is capable of directly *quantifying* the rates of forgetting (and degrees of learning) for any number of conditions. Deriving this quantity allows one to compare with greater precision the rates of forgetting between two conditions, and also makes it possible to generate reasonable predictions about where future memory performance will lie.

*Implications for the Testing Effect*

In the current study, it is clear that testing produced better overall recall than restudying. One reason for this benefit could be that the act of recall per se is more beneficial than restudying the material over again. Many past studies have shown that tests without feedback are significantly more beneficial than restudy opportunities, at least when the information is recalled correctly on the intervening test (e.g., Carpenter & DeLosh, 2005, 2006; Kuo & Hirshman, 1996, 1997; Roediger & Karpicke, 2006a, 2006b). In the case of tests with feedback, another reason why tests may be beneficial is because they reveal which items have been sufficiently learned and which ones require further study. For example, after trying to recall something and failing, subjects may, in subsequent study opportunities (e.g., when correct answer feedback becomes available), find new and better ways of encoding the information (e.g., see Izawa, 1992; LaPorte & Voss, 1975).

The tendency for tests to reduce forgetting does not appear to be as strong as that found in some prior studies (e.g., Roediger & Karpicke, 2006a; Thompson et al., 1978; Wenger et al., 1980; Wheeler et al., 2003). These prior studies utilized the ANOVA-based approach and sometimes observed a cross-over interaction between testing and retention interval. More specifically, restudying was sometimes more effective than testing at short retention intervals of a few minutes, but testing was more effective than restudying at longer retention intervals of at least one day. Unlike these studies, we found that testing was more effective than restudying at our five-minute retention interval, and so we did not observe the same type of cross-over interaction.

If we had observed the type of cross-over interaction that others have reported, then we certainly would have concluded that tests slow down the rate of forgetting according to the ANOVA-based approach. We also would

have found a stronger tendency for tests to slow down forgetting according to the curve-fitting approach. As we discussed previously, the curve-fitting approach can conclude that the rate of forgetting is lower for Test/Study than for Study even when the Test x Retention Interval Interaction is not significant and the two curves appear to be parallel. When there is a significant cross-over interaction between test and retention interval such that the two curves “diverge,” however, the curve-fitting approach would detect an even stronger reduction in forgetting for Test/Study vs. Study.

Why did we not observe this type of cross-over interaction, whereas past studies did? One point on which our study differs from these past studies is the use of feedback, and that could be one factor that influences the strength of this interaction. When tests are not accompanied by feedback, some items (i.e., those that were not correctly retrieved) might not benefit from testing (e.g., see Pashler et al., 2005). Indeed, past studies using tests without feedback have reported that after brief intervals of up to 10 minutes, testing was not any more beneficial than restudying (Slamecka & Katsaiti, 1988; Wenger et al., 1980). However, Thompson et al. (1978) found that this trend reversed when feedback was provided during the test. Without feedback, free recall on a final test delayed by 20 minutes was better for items that had been learned through restudying than through testing. With feedback, however, final test performance was better for items that had been learned through testing than through restudying. Consistent with Thompson et al. is the present finding that tests with feedback are more beneficial to learning than restudy opportunities even after a brief delay of only five minutes. Future research would benefit by exploring other factors that might influence the interaction between test condition and retention interval.

Another important aspect of the present study is that a testing advantage was produced without any overt responses having been elicited from the subject. During testing, subjects were instructed to try to recall the answer to each fact covertly (i.e., without saying the answer aloud, or typing it in). The significant effects of tests on learning therefore appear to be driven by the inward act of retrieval, and not the outward production of a response (e.g., see also Carpenter et al., 2006; Carpenter & Pashler, 2007). These data encourage the notion that tests can be used to improve learning in contexts in which it is not possible or convenient to collect and score overt responses (e.g., spatial and perceptual memory tasks such as face recognition).

These results could help shed some light on the nature of the testing effect and its utility to enhance learning and retention in practical settings. The extent to which testing reflects an enhancement in the degree of learning and a reduction in the rate of forgetting offers useful information for evaluating hypotheses that have been proposed to explain the effect (e.g., see Carrier & Pashler, 1992; Carpenter & DeLosh, 2006; Carpenter et al., 2006; Jacoby, 1978; Mozer, Howe, & Pashler, 2004), and in guiding future theoretical work.

These results also help to inform researchers and educators about the practical benefits of testing over time. When information has been tested, that information will be remembered better over time, compared to having restudied the information. This test-induced benefit is apparently stronger after providing repeated tests over the same information. These results suggest that tests should often be utilized in educational contexts to maximize retention of information over long time periods.

## References

- Allen, G. A., Mahler, W. A., & Estes, W. K. (1969). Effects of recall tests on long-term retention of paired associates. *Journal of Verbal Learning & Verbal Behavior*, 8, 463-470.
- Anderson, J. R. (2000). *Learning and memory: An integrated approach*. New York: John Wiley & Sons, Inc.
- Birnbaum, M. (1999). Testing critical properties of decision making on the Internet. *Psychological Science*, 10, 399-407.
- Bjork, R. A. (1988). Retrieval practice and the maintenance of knowledge. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 397-401). New York: Academic Press.
- Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name-learning. *Applied Cognitive Psychology*, 19, 619-636.
- Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition*, 34, 268-276.
- Carpenter, S. K., & Pashler, H. (2007). Testing beyond words: Using tests to enhance visuospatial map learning. *Psychonomic Bulletin & Review*, 14, 474-478.
- Carpenter, S. K., Pashler, H., & Vul, E. (2006). What types of learning are enhanced by a cued recall test? *Psychonomic Bulletin & Review*, 13, 826-830.
- Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, 20, 633-642.

- Chan, J. C. K., McDermott, K. B., & Roediger, H. L. III. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*, 135, 553-571.
- Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology*, 14, 215-235.
- Dempster, F. N. (1989). Spacing effects and their implications for theory and practice. *Educational Psychology Review*, 4, 309-330.
- Dempster, F. N. (1996). Distributing and managing the conditions of encoding and practice. In E. L. Bjork & R. A. Bjork (Eds.), *Handbook of perception and cognition: Memory*. San Diego, CA: Academic Press.
- Ebbinghaus, H. (1913). *Memory* (H. A. Ruger & C. E. Bussenius, Trans.). New York: Teachers College, Columbia University. (Original work published 1885).
- Glover, J. A. (1989). The “testing” phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, 81, 392-399.
- Izawa, C. (1992). Test trials contributions to optimization of learning processes: Study/test trials interactions. In A. F. Healy & S. M. Kosslyn (Eds.), *Essays in honor of William K. Estes, Vol. 1: From learning theory to connectionist theory* (pp. 1-33). Hillsdale, NJ: Erlbaum.
- Izawa, C., Maxwell, S., Hayden, R. G., Matrana, M., & Izawa-Hayden, A. J. E. K. (2005). Optimal foreign language learning and retention: Theoretical and applied investigations on the effects of presentation repetition programs. In C. Izawa & N. Ohta (Eds.), *Human learning and memory: Advances in theory and application: The fourth Tsukuba international conference on memory* (pp. 107-134). Mahwah, NJ: Erlbaum.
- Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior*, 17, 649-667.
- Krantz, J. H., & Dalal, R. (2000). Validity of web-based psychological research. In M. Birnbaum (Ed.), *Psychological experiments on the internet* (pp. 35-60). San Diego, CA: Academic Press.
- Kuo, T. M., & Hirshman, E. (1996). Investigations of the testing effect. *American Journal of Psychology*, 109, 451-464.
- Kuo, T. M., & Hirshman, E. (1997). The role of distinctive perceptual information in memory: Studies of the testing effect. *Journal of Memory & Language*, 36, 188-201.
- Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625-632). New York: Academic Press.
- LaPorte, R. E., & Voss, J. F. (1975). Retention of prose materials as a function of postacquisition testing. *Journal of Educational Psychology*, 67, 259-266.
- Loftus, G. R. (1985). Evaluating forgetting curves. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 11, 397-406.

- Loftus, G. R. (2002). Analysis, interpretation, and visual presentation of experimental data. In H. Pashler & J. T. Wixted (Eds.), *Steven's handbook of experimental psychology* (pp. 339-390). New York: John Wiley & Sons.
- McDaniel, M. A., & Fisher, R. P. (1991). Tests and test feedback as learning sources. *Contemporary Educational Psychology*, 16, 192-201.
- McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (in press). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*.
- McGraw, T. O., Tew, M. D., & Williams, J. E. (2000). The integrity of web-delivered experiments: Can you trust the data? *Psychological Science*, 11, 502-506.
- Modigliani, V. (1976). Effects on a later recall by delaying initial recall. *Journal of Experimental Psychology: Human Learning & Memory*, 2, 609-622.
- Mozer, M. C., Howe, M., & Pashler, H. (2004). Using testing to enhance learning: A comparison of two hypotheses. *Proceedings of the Twenty Sixth Annual Conference of the Cognitive Science Society* (pp. 975-980). Hillsdale, NJ: Erlbaum Associates.
- Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, 47, 90-100.
- Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 31, 3-8.
- Pashler, H., Rohrer, D., Cepeda, N. J., & Carpenter, S. K. (2007). Enhancing learning and retarding forgetting: Choices and consequences. *Psychonomic Bulletin & Review*, 14, 187-193.
- Postman, L., & Phillips, L. W. (1961). Studies in incidental learning: A comparison of the methods of successive and single recalls. *Journal of Experimental Psychology*, 61, 236-241.
- Raffel, G. (1934). The effect of recall on forgetting. *Journal of Experimental Psychology*, 17, 828-838.
- Reips, U-D. (2002). Standards for Internet experimenting. *Experimental Psychology*, 49, 243-256.
- Rickard, T. C. (2004). Strategy execution in cognitive skill learning: An item-level test of candidate models. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 30, 65-82.
- Roediger, H. L. III., & Karpicke, J. D. (2006a). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17, 249-255.
- Roediger, H. L. III., & 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. III., & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 31, 1155-1159.
- Rubin, D. C., & Wenzel, A. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review*, 103, 734-760.
- Runquist, W. (1983). Some effects of remembering on forgetting. *Memory & Cognition*, 11, 641-650.
- Runquist, W. (1986a). Changes in the rate of forgetting produced by recall tests. *Canadian Journal of Psychology*, 40, 282-289.
- Runquist, W. (1986b). The effect of testing on the forgetting of related and unrelated associates. *Canadian Journal of Psychology*, 40, 65-76.

- Runquist, W. (1987). Retrieval specificity and the attenuation of forgetting by testing. *Canadian Journal of Psychology*, 41, 84-90.
- Slamecka, N. J., & Katsaiti, L. T. (1988). Normal forgetting of verbal lists as a function of prior testing. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 14, 716-727.
- Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology*, 30, 641-656.
- Thompson, C. P., Wenger, S. K., & Bartling, C. A. (1978). How recall facilitates subsequent recall: A reappraisal. *Journal of Experimental Psychology: Human Learning & Memory*, 4, 210-221.
- Vul, E., & Pashler, H. (2007). Is incubation helpful only when people are misdirected? *Memory & Cognition*, 35, 701-710.
- Wenger, S. K., Thompson, C. P., & Bartling, C. A. (1980). Recall facilitates subsequent recognition. *Journal of Experimental Psychology: Human Learning & Memory*, 6, 135-144.
- Wheeler, M. A., Ewers, M., & Buonanno, J. F. (2003). Different rates of forgetting following study versus test trials. *Memory*, 11, 571-580.
- Wheeler, M. A., & Roediger, H. L. III. (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science*, 3, 240-245.
- Wickelgren, W. A. (1974). Single-trace fragility theory of memory dynamics. *Memory & Cognition*, 2, 775-780.
- Wilson, M. D. (1988). The MRC psycholinguistic database: Machine readable dictionary, version 2. *Behavioral Research Methods, Instruments & Computers*, 20, 6-11.
- Wixted, J. T. (1990). Analyzing the empirical course of forgetting. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 16, 927-935.
- Wixted, J. T. (2004). On common ground: Jost's (1897) law of forgetting and Ribot's (1881) law of retrograde amnesia. *Psychological Review*, 111, 864-897.
- Wixted, J. T., & Carpenter, S. K. (2007). The Wickelgren Power Law and the Ebbinghaus Savings Function. *Psychological Science*, 18, 133-134.
- Wixted, J. T., & Ebbesen, E. B. (1991). On the form of forgetting. *Psychological Science*, 2, 409-415.
- Wixted, J. T. & Ebbesen, E. B. (1997). Genuine power curves in forgetting: A quantitative analysis of individual subject forgetting functions. *Memory & Cognition*, 25, 731-739.
- Yale University. (2005). *The Kamusi project internet living Swahili dictionary*. Retrieved March 25, 2005, from <http://www.yale.edu/swahili/home.htm>

## Notes

<sup>1</sup> A scaling constant is needed because time is measured in arbitrary units. We make the simplifying assumption that the value of this parameter does not differ across subjects or conditions (see Wixted & Carpenter, 2007).

<sup>2</sup> This design feature made it possible for subjects to vary the functional presentation time of each item, so we recorded the amount of time it took for subjects to press "continue" after each item in both conditions. We refer to these response times as the *post-Test/Study ISI* and the *post-Study ISI*, respectively. In all three experiments, the *post-Test/Study ISI* did not differ significantly from the *post-Study ISI*. The *post-Test/Study ISI* and *post-Study ISI* were, respectively, 1567.11 ( $SD = 1569.11$ ) vs. 1573.38 ( $SD = 1857.40$ ) for Experiment 1, 1655.51 ( $SD = 1483.81$ ) vs. 1859.65 ( $SD = 2192.46$ ) for Experiment 2, and 2736.50 ( $SD = 3521.62$ ) vs. 3087.25 ( $SD = 4140.62$ ) for Experiment 3 (all  $ps > .05$ ).

<sup>3</sup> Paired samples t-tests could also be used to examine differences between the means of the degree of learning and rate of forgetting parameters for Test/Study vs. Study. For a few subjects, however, the parameter estimates from the curve-fitting procedure varied to an extreme degree from those produced by the large majority of subjects. The presence of these outlier scores suggested that the population of parameter estimates from which our sample was drawn may not be normally-distributed. We therefore analyzed these data using the non-parametric sign test, which, unlike the t-test, does not rely on the assumption that the population be normally-distributed. Results of the t-test support the conclusion that Test/Study increases the degree of learning and reduces the rate of forgetting more than Study, but these effects were significant only when the outlying values were excluded. Whether or not these outliers were included did not affect the results of the ANOVA-based analysis.

<sup>4</sup> The only significant effect of environment occurred in the form of a small 3-way interaction between test condition, retention interval, and environment during Sessions 1 through 4. Rather than a reflection of any effect of environment on performance, however, this interaction was most likely a statistical artifact of small and

unequal sample sizes across the different environmental conditions. It is common for individual subjects to vary in their patterns of forgetting across conditions (e.g., some subjects may forget little across the first few sessions for Test/Study and forget more for Study, other subjects may show a similar pattern of decline for Test/Study and Study), and this variability can create such interactions when subjects are few and unevenly distributed across environments. Subjects across all environments showed the same general pattern in which performance decreased over time and Test/Study outperformed Study.

#### ARCHIVED MATERIALS

The following materials and links may be accessed through the Psychonomic Society's Norms, Stimuli, and Data Archive, <http://www.psychonomic.org/archive>. To access these files or links, search the archive for this article using the journal (Memory and Cognition), the first author's name (Carpenter) and the publication year (2007).

FILE: Carpenter-MC-2007.zip

DESCRIPTION: The compressed archive file contains two files: carpenter2007exp1exp2.xls, containing the 60 obscure facts used in Experiments 1 and 2, and carpenter2007exp3.xls, containing the 60 Swahili-English word pairs used in Experiment 3.