

# The effects of tests on learning and forgetting

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

In three experiments, we investigated whether memory tests enhance learning and reduce forgetting more than additional study opportunities do. 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 5 min or 1, 2, 7, 14, or 42 days. We explored forgetting by means of an ANOVA and also by fitting a power function to the data. In all three experiments, testing enhanced overall recall more than restudying did. According to the power function, in two out of three experiments, testing also reduced forgetting more than restudying did, although this was not always the case according to the ANOVA. We discuss the implications of these results both for 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 [www.psychonomic.org/archive](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, 2006a; Wheeler & Roediger, 1992). It is often found that tests strengthen memory even more than do 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 either 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 than for words learned through additional study. The beneficial effect of testing versus restudying—that is, the *testing effect*—has 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).

Given the potential of tests to enhance learning, a number of psychologists have argued that tests should be used frequently in educational contexts not merely to assess learning, as is the standard practice, but to promote 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 (2007) found that students enrolled in an online brain and behavior course performed significantly better on the final exam when they reviewed the course information by taking quizzes rather than by doing additional reading.

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

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

Some researchers have explored this issue by comparing the effects of testing versus restudying on two different tests: an immediate test and a test that is delayed by an interval of up to 1 week. In four studies, restudying was as beneficial as or more beneficial than testing when retention was measured after several minutes, but testing was more beneficial than restudying when retention was measured

after 2 days (Thompson, Wenger, & Bartling, 1978; Wenger, Thompson, & Bartling, 1980) or after 7 days (Roediger & Karpicke, 2006b; Wheeler, Ewers, & Buonanno, 2003). On the basis of 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 (see, e.g., Wheeler et al., 2003).

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 of items in a test condition with retention of items in a restudy condition. Instead, the studies compared retention of items in a test condition without feedback with retention of other items in a no-test control condition. As such, it is possible that any apparent differences in the rate of forgetting reflected differences in the amount of overall exposure time rather than the effects of 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 assessed on the basis of the number of items 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 only for the easier and/or initially better-learned items. Therefore, even if tests did 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.

One study known to the present authors 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 1-day delay, or after a 5-day delay. Across the 1- and 5-day-delayed tests, there was no significant advantage for testing versus restudying and no interaction between test condition and retention interval. These results are difficult to interpret, however, since, unlike previous researchers 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 does appear to require further exploration. Although some studies report that tests appear to slow down forgetting for up to 7 days (Roediger & Karpicke, 2006b; 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 and no re-exposure, which is quite different from comparisons of testing and restudying. The present study set out to resolve what might be considered equivocal evidence about the effect of testing on forgetting for a period of up to several days and to provide much-needed data on the effects of

testing versus those of restudying on forgetting over even longer time periods.

### The Present Study

The present study explored the effects of tests on learning and forgetting over a much longer range of time than has ever been explored in past research. Whereas past studies measured retention as a function of testing versus restudy opportunities at two or three points in a time span of up to 1 week, our study explored memory at six different points in a time span of 6 weeks. We compared retention of a group of items that was learned through an additional study opportunity with retention of 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 5 min, 1 day, 2 days, 7 days, 14 days, or 42 days. We used tests of cued recall in order to insure that individual items could be equated according to the amount of study time that 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, they were 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 on tests 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); likewise, on cued recall tasks, items that elicit an error and are 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 4 sec to recall the answer to each question about a fact presented 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 an additional 2 sec. For the restudy condition (referred to as *study*), subjects were presented with both the question and the answer (e.g., “What breed of dog has the best eyesight? Greyhound”) for 6 sec. In the test/study condition of Experiment 3, subjects were given 2 sec 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 an additional 2 sec. In the study condition of Experiment 3, subjects were shown a 4-sec presentation of the complete word pair (“somo”–“friend”). Thus, our design, which follows that of Carrier and Pashler (1992), ensured that the amount of time subjects spent on each trial in the test/study condition equaled the amount of time that they spent on each trial in the study condition. Furthermore, it is worth noting that the test/study procedure actually involved a reduced amount of time during which the correct answer was presented, rela-

tive to the study procedure. 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 opportunity on each fact. Past research has shown that the effects of tests are stronger when the number of tests and 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 opportunities or three study opportunities for each fact. Experiment 3 also provided subjects with three test/study or three study opportunities for 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 ANOVA to compare forgetting rates. In these studies, 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 the percentage of correct answers 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 is a substantial research tradition on this topic, beginning with Ebbinghaus (1885/1913), who first showed that the decline in percentage 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 across different species of subjects (see, e.g., Wickelgren, 1974; Wixted, 2004).

The power function analyzes forgetting using an approach different from that used by an 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 versus 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 for measuring the time course of forgetting.

Our study addressed the following questions: After sampling recall across a 6-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? Next, (3) are the effects of tests strengthened by providing three test/study versus three study opportunities, as opposed to providing only one test/study versus one study opportunity? 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

### Method

**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 it 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 (see, e.g., Birnbaum, 1999; Krantz & Dalal, 2000; McGraw, Tew, & Williams, 2000; Reips, 2002), and it is our impression that our paid Internet panelists in particular are generally more careful and attentive than are subjects drawn from the typical undergraduate subject pool. Subjects performed six sessions (the first was about 20–35 min in length, and the remaining five sessions lasted about 2 min each) in exchange for payment of \$20.00.

**Materials and Design.** We used a variety of online 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 [www.psychonomic.org/archive](http://www.psychonomic.org/archive).

We used a  $2 \times 6$  (test condition: test/study or study  $\times$  retention interval: 5 min or 1, 2, 7, 14, or 42 days) within-subjects design. For each subject, 5 facts were randomly assigned to one of the 12 possible conditions. To measure recall across time, a different group of 10 facts (5 from test/study and 5 from study) was tested after 5 min or 1, 2, 7, 14, or 42 days.

**Procedure.** For all data collection, we used a Web site that ran on the free and open source LAMP (Linux, Apache, MySQL, and PHP) framework. This Web site was tested in order to ensure its accessibility from 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 gender, age, level of education, and in what type of environment they would complete 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, that told them

that they would be learning obscure facts and that 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 6 sec. 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 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 the test/study condition, subjects were presented with the fact in question format and were instructed to covertly recall the correct one-word answer within 4 sec. After 4 sec, the correct answer was displayed, along with the question, for 2 additional sec. Thus, during test/study trials, subjects were not required to enter a response but instead simply required to recall the answer in their minds. Therefore, any benefits conferred by the test/study procedure cannot be attributed to the overt response (e.g., typing in or writing down an answer) but rather must be attributed to the act of recalling the response and then receiving feedback. For the study condition, subjects were presented with the fact in question format and the correct answer for 6 sec. Thus, the total time that each item was presented in both the test/study and study conditions was always 6 sec. 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 a 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 5-min video game distractor task.

Immediately following the distractor task, subjects were given a final test over 10 of the facts (5 from test/study and 5 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 2–6, subjects were given the same type of test again, but over 10 different facts (5 from test/study and 5 from study).

When the time came for a subject to perform Sessions 2–6, 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 h following Session 1; Session 3 could be completed between 42 and 56 h following Session 1; Session 4, between 156 and 192 h following Session 1; Session 5, between 312 and 384 h following Session 1; and Session 6, between 984 and 1,080 h following Session 1. A new group of 10 facts (5 from test/study and 5 from study) was tested in each of Sessions 2–6, and each of these sessions lasted approximately 2 min. The total time to complete all 6 sessions was approximately 25 min in Experiment 1 and 45 min in Experiment 2.

## Results and Discussion

**Effects of environment.** The majority of subjects participated while in a room by themselves—at least 67% in each of the six sessions of 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  $2 \times 6$  (test condition  $\times$  retention interval) repeated measures ANOVA. According to the ANOVA, the main effect of test condition was significant in Experiment 1 [ $F(1,54) = 14.36, p < .001, MS_e = .032, \eta_p^2 = .21$ ] and in Experiment 2 [ $F(1,56) = 47.38, p < .001, MS_e = .023, \eta_p^2 = .46$ ]. 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 conditions were retained at an average overall rate of 78%, whereas facts learned through three study conditions were retained at an average overall rate of 70%.

In both experiments, significant forgetting occurred across the 6-week time interval. The main effect of retention interval was significant in Experiment 1 [ $F(5,270) = 168.73, p < .001, MS_e = .038, \eta_p^2 = .76$ ] and in Experiment 2 [ $F(5,280) = 147.51, p < .001, MS_e = .036, \eta_p^2 = .72$ ]. The test condition  $\times$  retention interval interaction was not significant in Experiment 1 ( $F < 1$ ); however, this interaction was significant in Experiment 2 [ $F(5,280) = 3.88, p < .01, MS_e = .023, \eta_p^2 = .06$ ]. The mean proportion of facts recalled in all conditions for Experiments 1 and 2 is 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 and retention interval made it possible for us to fit the power function to the data for each individual subject (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 (see, e.g., 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 proportion of correctly recalled facts across the six retention intervals for Experiment 1 (panel A) and Experiment 2 (panel B). The smooth curves represent the average of the 55 individual subjects' forgetting curves for Experiment 1, and the average of the 57 individual subjects' forgetting curves for Experiment 2.

We used a binomial sign test to evaluate whether a significant number of subjects exhibited higher degrees of learning and lower rates of forgetting for test/study than for study.<sup>3</sup> In Experiment 1, test/study showed an advan-

**Table 1**  
**Mean Proportion of Items Retained (With Standard Errors)**  
**As a Function of Retention Interval and Test Condition**

	Retention Interval													
	5 min		1 day		2 days		7 days		14 days		42 days		Total	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Experiment 1														
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														
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														
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		

Note—*Test/study* refers to test trials with feedback; *Study* refers to pure study trials. In Experiment 1, subjects had one test/study or one study session over obscure facts; in Experiment 2, they had three test/study or three study sessions over obscure facts; and in Experiment 3, subjects had three test/study or three study sessions over Swahili–English word pairs.

tage in the degree of learning for the majority of subjects (60%) compared with study, and 60% of the subjects also showed a lower rate of forgetting in the test/study condition than in the study condition. 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 in the test/study condition than in the study condition ( $p < .01$ ), and a significant proportion of subjects (68%) showed lower rates of forgetting for test/study than for study ( $p < .01$ ).

The apparent difference in the rate of forgetting between test/study and 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 5-min 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 5-min retention interval.

### EXPERIMENT 3

#### Method

**Subjects.** Forty-four subjects, drawn from the same pool used in Experiments 1 and 2, completed Experiment 3. None of these subjects had participated in either of the two previous experiments. Subjects completed six sessions (the first was about 15–20 min in length, and the remaining five sessions lasted about 2 min each), and were paid \$10.00. Subjects (35 females and 9 males) ranged in age from 19–63 years, 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 and one and three syllables in length,

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 the Kamusi Project Web site (Yale University, 2005).

**Design and Procedure.** All aspects of the design were identical to those of 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 4 sec. In the test/study condition, subjects saw only the Swahili word (“somo”) for 2 sec and were instructed to try to covertly recall the correct English translation (“friend”). After 2 sec had elapsed, the complete word pair was presented (“somo”–“friend”) for 2 additional seconds. In the study condition, subjects saw the complete word pair for a total of 4 sec.

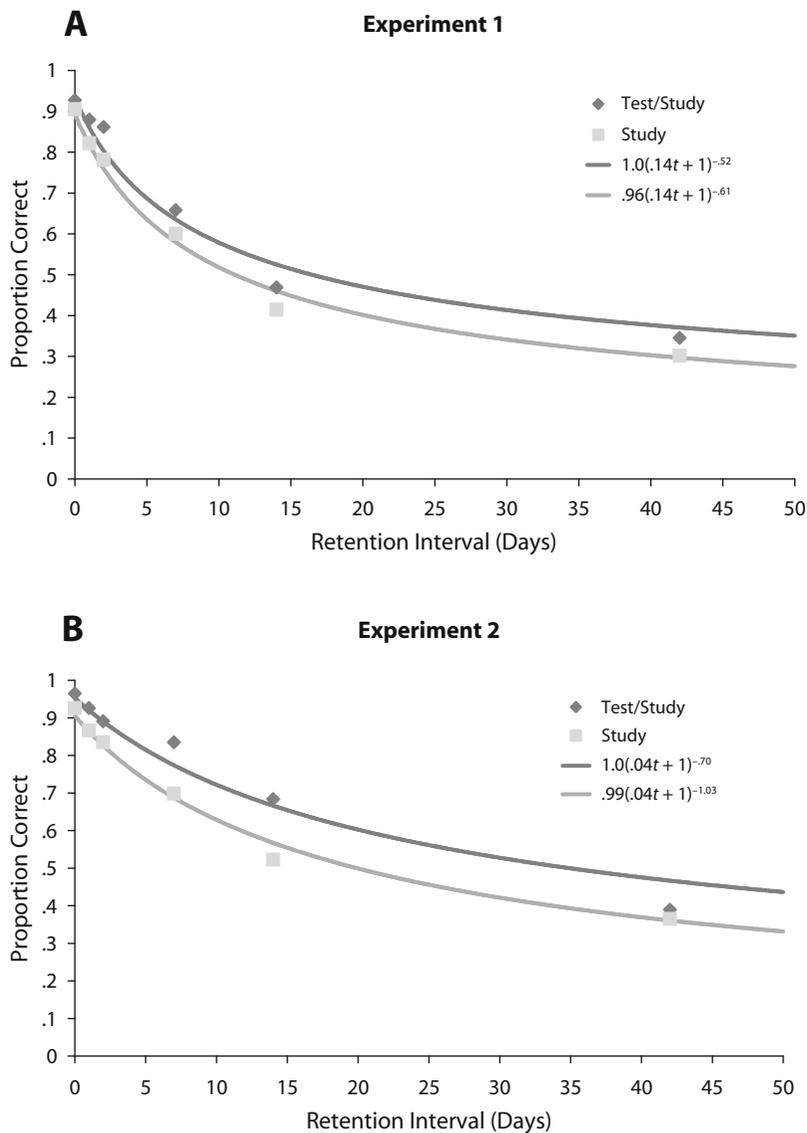
#### Results and Discussion

**Effects of environment.** At each of the six sessions in Experiment 3, at least 66% of the 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 versus study.** As in the previous experiments, we analyzed the data from Experiment 3 using the ANOVA-based method and the curve-fitting method. According to the ANOVA, the main effect of test condition was significant [ $F(1,43) = 30.22$ ,  $p < .001$ ,  $MS_e = .023$ ,  $\eta_p^2 = .41$ ], as was the main effect of retention interval [ $F(5,215) = 28.41$ ,  $p < .001$ ,  $MS_e = .057$ ,  $\eta_p^2 = .40$ ]. 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 6-week time interval. However, the test condition  $\times$  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 proportion of correctly recalled words across the six retention intervals for Experiment 3. The smooth curves represent the average of the 44 individual subjects’ forgetting curves.

Of the 44 subjects, 27 exhibited a higher degree of learning in the test/study condition than in the study con-



**Figure 1.** Subjects were given a test with feedback (*test/study*) or a restudy opportunity (*study*) for each fact. Recall of these facts was tested after 5 min or 1, 2, 7, 14, or 42 days; in Experiment 1, recall was tested following just one test/study or one study opportunity (results shown in panel A); in Experiment 2, it was tested following three test/study or three study opportunities (results shown in panel B). The points represent the average proportion of facts recalled from test/study versus 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 opportunity increased the degree of learning and reduced the rate of forgetting over having just one study opportunity (A), but these effects did not reach significance. Having three test/study opportunities significantly increased the degree of learning and significantly reduced the rate of forgetting over having three study opportunities (B). The smooth curves represent the mean of the 55 individual subjects' forgetting curves in Experiment 1 and the 57 individual subjects' 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.

dition, 15 exhibited a higher degree of learning in the study condition, and 2 exhibited the same degree of learning in both the test/study and the study conditions. The proportion of subjects showing higher degrees of learning

for test/study (64%) approached significance ( $p = .09$ ). Of the 44 subjects, 29 exhibited a lower rate of forgetting in the test/study condition than in the study condition, 13 exhibited a lower rate of forgetting in the study condition,

and 2 exhibited the same rate of forgetting in both the test/study and the study conditions. The proportion of subjects showing lower rates of forgetting in test/study (69%) than in study was significant ( $p < .05$ ).

## GENERAL DISCUSSION

In three experiments, the test/study condition was more beneficial to memory recall than was the study condition. We observed that one test/study opportunity provided a significant benefit for the recall of obscure facts as compared with one study opportunity (Experiment 1) and that three test/study opportunities provided an apparently larger benefit as compared with three study opportunities (Experiment 2). Furthermore, we observed the same benefit from three test/study opportunities versus three study opportunities on the more difficult recall of Swahili–English word pairs (Experiment 3). These results are consistent with a number of prior studies showing that, as assessed through cued recall, testing enhances memory more than restudying does (see, e.g., Carpenter & DeLosh, 2005, 2006; Carpenter et al., 2006; Carrier & Pashler, 1992; Izawa, 1992).

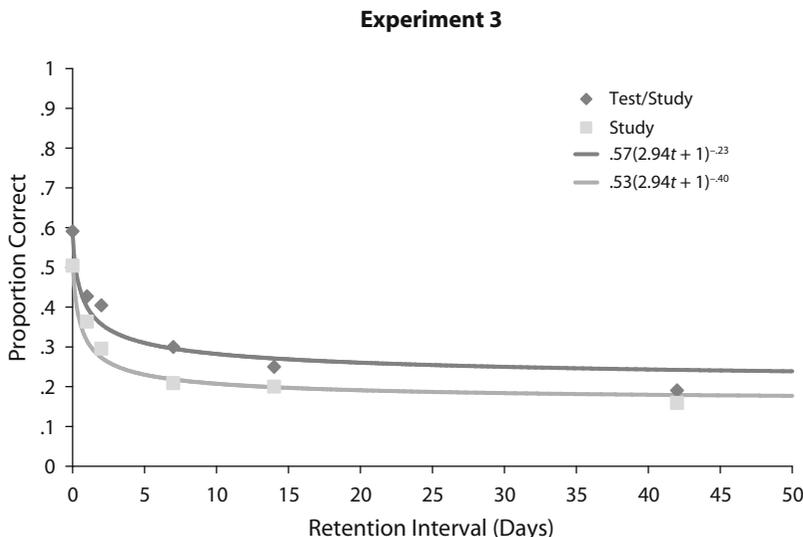
This finding was observed despite the fact that the total study time in the test/study and study conditions was equal. Indeed, the advantage of the test/study procedure was actually produced by depriving the subject of certain information (i.e., the one-word answer to the fact query or the correct English word for the Swahili cue) for 2 sec, meaning that the amount of study time that the subjects had 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 opportu-

nities, 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 (see, e.g., Roediger & Karpicke, 2006b; 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  $\times$  retention interval interaction and therefore concluded that tests do not slow down the rate of forgetting. The conclusions drawn from the present data depend on which approach one takes to measuring the rate of forgetting.

### Different Approaches to Measuring Forgetting

In this article, 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  $\times$  retention interval interaction was not significant, suggesting that the rate of forgetting for the test/study condition did not differ from that for the study condition. According to the curve-fitting approach, however, the rate of forgetting was consistently lower for the test/study than for the study condition. Although this difference did not reach signifi-



**Figure 2.** Experiment 3 replicated Experiment 2 but used Swahili–English word pairs instead of obscure facts. Having three test/study opportunities for Swahili–English word pairs enhanced the degree of learning compared with having three study opportunities, but this effect was not significant. Having three test/study opportunities for Swahili–English word pairs significantly reduced the rate of forgetting compared with having three study opportunities. The smooth curves represent the mean of the 44 individual subjects' forgetting curves.

cance 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 according to the ANOVA-based analysis, this is not always the case? A nonsignificant 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 versus 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 toward an asymptote greater than zero; that is, even if the lower curve (study) projects toward an asymptote of zero, the upper curve (test/study) would have to project toward 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 nonsignificant 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 toward an asymptote of zero. If this is true (see, e.g., 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 (see, e.g., 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 toward zero, the test/study curve takes longer than the study curve to fall below any given level of performance. To fall below 25%, for example, the curve for study requires about 3 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 and the curve-fitting approaches 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 Versus Using the 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 (e.g., in the present study, only five items in each of the 12 within-subjects conditions), which can increase the chances that one item will have a nontrivial 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, since it requires 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 nonnormal 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 nonparametric 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 that subjects in Condition A initially recall 22% of the items on a word list, whereas subjects in Condition B recall 90% of them. Suppose that after 1 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 toward different asymptotes). However, an absolute loss of 20% is a far smaller proportion of the material originally learned in Condition B than in 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 toward an asymptote of zero).

The power function nicely characterizes the curvilinear form of forgetting, whereby the amount of information lost from memory is greatest at first and then gradually decreases 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 (see, 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 the 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 present study, it is clear that testing produced better overall recall than did restudying. One reason for this benefit could be that the act of recall per se is more beneficial than studying the material 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 (see, e.g., Carpenter & DeLosh, 2005, 2006; Kuo & Hirshman, 1996, 1997; Roediger & Karpicke, 2006a, 2006b). In the case of tests with feedback, another reason that tests may be beneficial is that 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 (see, e.g., Izawa, 1992; LaPorte & Voss, 1975).

The tendency found in the present study for tests to reduce forgetting does not appear to be as strong as that found in some prior studies (see, e.g., Roediger & Karpicke, 2006b; Thompson et al., 1978; Wenger et al., 1980; Wheeler et al., 2003). These prior studies utilized the ANOVA-based approach and sometimes observed a crossover 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 1 day. In contrast to these findings, we found that testing was more effective than restudying at our 5-min retention interval, and thus we did not observe the same type of crossover interaction.

If we had observed the type of crossover interaction that others have reported, then we certainly would have concluded that, according to the ANOVA-based approach, tests slow down the rate of forgetting. 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  $\times$  retention interval interaction is not significant, and the two curves appear to be parallel. When there is a significant crossover 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 than for study.

Why did we not observe this type of crossover interaction, whereas past researchers did? One point on which our study differs from these past studies is the use of feedback, which 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 (see, e.g., Pashler et al., 2005). Indeed, past studies using tests without feedback have reported that after brief intervals of up to 10 min, test-

ing 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 min 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 are restudy opportunities even after a brief delay of only 5 min. Future research would benefit from 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 query 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 (see, e.g., Carpenter & Pashler, 2007; Carpenter et al., 2006). 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 ability 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 (see, e.g., Carpenter & DeLosh, 2006; Carpenter et al., 2006; Carrier & Pashler, 1992; Jacoby, 1978; Mozer, Howe, & Pashler, 2004) and for guiding future theoretical work.

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

### AUTHOR NOTE

E.V. is now with the Department of Brain and Cognitive Sciences at the 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, the 2006 Annual Meeting of the American Educational Research Association, and the 2007 All Hands Meeting of the Temporal Dynamics of Learning Center. We thank Brian Ilagan for his assistance in collecting obscure facts. Correspondence concerning this article should be addressed to S. K. Carpenter, Department of Psychology, 0109, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0109 (e-mail: scarpenter@ucsd.edu).

## 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: Wiley.
- BIRNBAUM, M. H. (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: Current research and issues* (Vol. 1, pp. 396-401). Chichester, U.K.: Wiley.
- 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*, **1**, 309-330.
- DEMPSTER, F. N. (1996). Distributing and managing the conditions of encoding and practice. In E. L. Bjork & R. A. Bjork (Eds.), *Memory* (Handbook of Perception and Cognition, 2nd ed., pp. 317-344). San Diego: Academic Press.
- EBBINGHAUS, H. (1913). *Memory* (H. A. Ruger & C. E. Bussenius, Trans.). New York: Columbia University, Teachers College. (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, & R. M. Shiffrin (Eds.), *Essays in honor of William K. Estes: Vol. 2. From learning processes to cognitive processes* (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* (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 & Verbal Behavior*, **17**, 649-667.
- KRANTZ, J. H., & DALAL, R. (2000). Validity of Web-based psychological research. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 35-60). San Diego: 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). London: 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. Wixted (Eds.), *Stevens' Handbook of experimental psychology: Vol. 4. Methodology in experimental psychology* (3rd ed., pp. 339-390). New York: Wiley.
- MCDANIEL, M. A., ANDERSON, J. L., DERBISH, M. H., & MORRISSETTE, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, **19**, 494-513.
- MCDANIEL, M. A., & FISHER, R. P. (1991). Tests and test feedback as learning sources. *Contemporary Educational Psychology*, **16**, 192-201.
- MCGRAW, K. 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. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society* (pp. 975-980). Mahwah, NJ: Erlbaum.
- 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-based 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). 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.
- 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. N. (1983). Some effects of remembering on forgetting. *Memory & Cognition*, **11**, 641-650.
- RUNQUIST, W. N. (1986a). Changes in the rate of forgetting produced by recall tests. *Canadian Journal of Psychology*, **40**, 282-289.
- RUNQUIST, W. N. (1986b). The effect of testing on the forgetting of related and unrelated associates. *Canadian Journal of Psychology*, **40**, 65-76.
- RUNQUIST, W. N. (1987). Retrieval specificity and the attenuation of forgetting by testing. *Canadian Journal of Psychology*, **41**, 84-90.
- SLAMECKA, N. J., & KATSARITIS, 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). Incubation benefits only after people have been 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. (1988). MRC Psycholinguistic Database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers*, **20**, 6-10.
- 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**, 864-879.
- 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 [www.yale.edu/swahili/home.htm](http://www.yale.edu/swahili/home.htm). Now available at [www.kamusiproject.org](http://www.kamusiproject.org).

#### NOTES

1. 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).

2. 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 the presentation of each item in both conditions. We refer to these response times as the *posttest/study interstimulus interval* (ISI) and the *poststudy ISI*, respectively. In all three experiments, the posttest/study ISI did not differ significantly from the poststudy ISI. The posttest/study ISI and the poststudy ISI were, respectively, 1,567.11 ( $SD = 1,569.11$ ) versus 1,573.38 msec ( $SD = 1,857.40$ ) for Experiment 1; 1,655.51 ( $SD = 1,483.81$ ) versus 1,859.65 msec ( $SD = 2,192.46$ ) for Experiment 2; and 2,736.50 ( $SD = 3,521.62$ ) versus 3,087.25 msec ( $SD = 4,140.62$ ) for Experiment 3 (all  $ps > .05$ ).

3. Paired-samples *t* tests could also be used to examine differences between the means of the degree-of-learning and rate-of-forgetting pa-

rameters for the test/study versus the study condition. 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 nonparametric sign test, which, unlike the *t* test, does not rely on the assumption that the population is 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 does 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.

4. The only significant effect of environment occurred in the form of a small three-way interaction between test condition, retention interval, and environment during Sessions 1-4. Rather than being 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, whereas 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 associated with this article may be accessed through the Psychonomic Society's Norms, Stimuli, and Data archive, [www.psychonomic.org/archive](http://www.psychonomic.org/archive).

To access these files, search the archive for this article using the journal name (*Memory & Cognition*), the first author's name (Carpenter), and the publication year (2008).

FILE: Carpenter-MC-2008.zip.

DESCRIPTION: The compressed archive file contains two files: carpenter2008exp1exp2.xls, containing the 60 obscure facts used in Experiments 1 and 2.

carpenter2008exp3.xls, containing the 60 Swahili-English word pairs used in Experiment 3.

AUTHOR'S E-MAIL ADDRESS: [scarpenter@ucsd.edu](mailto:scarpenter@ucsd.edu).

(Manuscript received January 16, 2007;  
revision accepted for publication September 3, 2007.)