Is Temporal Spacing of Tests Helpful Even When It Inflates Error Rates?

Harold Pashler
University of California, San Diego

Gregory Zarow
EmpiricalMethods

Baylor Triplett
University of California, San Diego

The occurrence of errors is often thought to impede associative learning. This was tested in 2 studies, each of which involved 2 sessions. In Session 1, subjects learned foreign vocabulary (Experiment 1) or obscure English words (Experiment 2) and received 2 tests (each with corrective feedback) separated by a variable lag. Greater lags drastically reduced performance on the 2nd test. However, they dramatically improved performance in a Session-2 test given 1 day (Experiment 1) or 1 week later (Experiment 2). This pattern held even for items that elicited errors on the 1st test of Day 1. Evidently, the benefit of spacing overwhelms any possible harmful effect of producing errors. The results have clear and nonobvious implications for computer-aided instruction.

Although many educators aspire to minimize the role of rote memorization in learning, the need for some memorization remains as great as ever. This need is seen, for example, in educational contexts ranging from second-language vocabulary learning to organic chemistry classes, from programming language instruction to the acquisition of basic historical facts. After a century of effort, research on learning and memory unfortunately appears to have provided little concrete guidance to help instructors and learners reduce the amount of time and effort required to accomplish rote learning or to retard the rate at which newly learned information is forgotten. Textbooks of instructional design typically do not even cite basic research in this area and offer little in the way of concrete guidance about how educational drills[1] should be arranged (Gagne, Briggs, & Wager, 1992). Present-day computer tutorial programs use a haphazard array of strategies for determining which items a learner should be studying next and for deciding when a given item has been adequately studied (Karlsson, 1989).

The goal of the present article is to shed light on a key question that arises whenever one must decide what information item someone should study next on a list of items to learn. The specific question considered here is: Should information items be drilled with short intervals separating the tests, so as to minimize errors and their possibly deleterious effects? Or should they be tested at longer intervals, to provide greater temporal distribution of practice (the *spacing effect*; Dempster, 1996)? When the subject makes an error on an item, should that item be retested sooner rather than later, perhaps in order to prevent errors from being inadvertently "stamped-in"?

This question holds substantial interest, both theoretical and practical, because two influential and intuitively compelling ideas clash. On the one hand, there is the idea of *error minimization,* first popularized by Skinner (1968). Skinner held that producing an error, even if it is followed by immediate correction, can impair learning. On the other hand, there is the spacing effect, often casually summarized as the dictum to use a long delay between study episodes. When study consists of drill, long delays are bound to make errors more frequent (as compared with shorter delays). Thus, both doctrines cannot be followed simultaneously. To minimize errors, one is led to drill an item at relatively short lags, especially if someone has made a mistake on that item (or to use study instead of drill, which has obvious disadvantages, as discussed in Footnote 1). The spacing principle, by contrast, seems to argue for maximizing the lag (Dempster, 1988, 1996). Before turning to the present study, we describe a bit more about each of these literatures.

[1] By drill, we mean iterative testing with feedback. We focus here on drill because testing with feedback has been found to be superior to simply restudying the same material over, especially with long retention intervals (Bjork, 1988; Carrier & Pashler, 1992; Cull, 2000; Izawa, 1970; Kuo & Hirshman, 1996). Additionally, testing allows the learner to concentrate on the most difficult items, and probably increases motivation as well. Thus, it is not surprising that an informal survey of commercially available language tutorial programs found that all used drill (N. Coburn, personal communication, June 1, 2002).

The Spacing Effect

The spacing effect is often summarized in the claim that greater temporal distribution of practice maximizes learning (see Crowder, 1976; Dempster, 1988, 1996, for reviews). This oversimplifies the matter, however. At least as far back as Austin (1921), it has been noted that when the final test occurs immediately after the completion of the final practice phase, spaced practice is often worse than massed practice. A number of more recent studies have shown that the advantage of spaced presentations is greater for longer retention intervals (Bahrick, Bahrick, Bahrick, & Bahrick, 1993; Glenberg, 1976; Glenberg & Lehman, 1980). Most studies of spacing have relied on pure study—simply presenting the material to be learned over and over to the subject—rather than on testing. However, a number of studies varied the spacing of paired-associate learning tests (with feedback) and found a beneficial effect of spacing on learning assessed at the conclusion of a given session (Bjork, 1966; Bregman & Wiener, 1970; Greeno, 1964; Rumelhart, 1967; Young, 1971).

Error Minimization

The idea that it may be wise to minimize errors during training also has a long pedigree. It may have originated with Pavlov's (1927) suggestion that when an animal errs in discrimination training, acquisition is profoundly delayed. This idea was one of the cornerstones of Edwin Guthrie's influential writings (e.g., Guthrie, 1952). Guthrie claimed that animals learn what they do (and especially what they do last) and consequently that making an error stamps in undesired stimulus–response associations. Guthrie implied that this happens even when people consciously recognize that they have erred. The assumption that errors have undesired consequences apparently underlies the advocacy of *error-free learning,* a prominent slogan of the Programmed Instruction movement that was influential in the 1960s (Taber, Glaser, & Schaefer, 1965; Vargas, 1986). Although programmed instruction seemed at least mildly effective (Kulik, Schwalb, & Kulik, 1982), it is not clear whether this should be attributed to the benefits of error minimization per se.

Laboratory evidence for the idea that errors have detrimental effects in associative learning is scarce and somewhat indirect. There is no doubt that when a subject makes an error on a particular item in a typical laboratory learning task, he or she will find this item relatively difficult to master. In recall of word lists, for example, when people make successive attempts to recall a list of previously studied words, failure to retrieve a given item in memory test n strongly predicts failure to retrieve the item in memory test $n + 1$ (Roediger & Payne, 1982). In cued recall, when people have studied an A–B association and erroneously retrieve response term B' given stimulus A on a trial, subsequent errors involving A tend to involve producing B' (Butler & Peterson, 1965). People also tend to repeat the same errors quite often when they are taught arbitrary discriminative responses to stimuli, even when immediate feedback is provided (Marx & Witter, 1972). Although they are intriguing, these findings have the limitations inherent in all correlational data. On one hand, producing an error on trial n might be producing deleterious learning, thus causing the same errors to recur on succeeding trials; on the other hand, errors might merely index the fact that a given item is relatively difficult for the learner. Thus, these data do not necessarily show that producing an error has a deleterious causal effect on later performance.

One compelling bit of evidence for the Guthrian idea comes from a lone, virtually uncited paired-associate learning study by Cunningham and Anderson (1968). They found that *forced confirmation* (requiring the subject to guess the response term on every trial, starting from the very first trial) markedly impaired learning. Because this procedure triggered many errors, a Guthrian interpretation seems plausible. However, the requirement to guess before having been given any information whatsoever seems so extreme that one should perhaps not conclude too much from this.

One intuitive interpretation is that it may be best to use fairly long spacings after a learner has responded correctly to a given item even once; however, whenever he or she makes an error, it may be best to test that item again very soon, so as to ensure that the error is not stamped in. This presents an empirical question and a possible opportunity to partially reconcile the spacing effect with the Guthrian interpretation of learning.

Present Study

The present study involved an initial learning session in which associated pairs were given an initial presentation followed by two tests (with feedback). The first goal of the present study was to assess how delaying the second of these tests would affect both performance during learning (as assessed on the same day) and the degree of learning (as reflected in a final test on a later day). The second goal was to see whether this dependency would change as a function of whether the person had succeeded or failed on the first test. Unlike the studies of spacing effects mentioned above, we chose materials that fairly closely reflect real associative-learning tasks people might undertake outside of the laboratory. We also gave a final test of learning after a meaningfully long delay (1 day and 1 week, respectively, in the two experiments) rather than merely measuring learning at the end of the learning session.

Experiment 1

Experiment 1 involved foreign language vocabulary learning. To make sure that subjects had little preexisting knowledge of what they would be taught, we used an Eskimo[2] language, primarily the Siberian Inupiaq dialect, that is rarely spoken on the campus of the University of California, San Diego. Subjects performed two sessions on consecutive days. On Day 1, subjects were exposed to 200 trials involving approximately 66 Eskimo–English word pairs, three times each. Exposure 1 was a study trial (the pair was simply presented). Exposures 2 and 3 were test trials, designated Test 1 and Test 2, in which the subject was shown the Eskimo word and asked to type in the English translation. If that response was incorrect, the subject was then shown the English translation. The nominal lag between the study trial and Test 1 (the first test trial) was 2 intervening items (lag = 2), chosen to ensure that subjects would make a significant number of errors on Test 1. The lag

---

[2] Unlike some Arctic peoples, the Inupiaq refer to themselves as Eskimos rather than Inuit (Sailer, 2002).

between Test 1 and Test 2 was assigned randomly (for reasons described below, nominal lags did not always correspond precisely to the actual lag). On Day 2, subjects were tested once without feedback on every word they had studied on Day 1.

Method

Subjects. Forty-three students from the University of California, San Diego, participated in two consecutive sessions 1 day apart. Two subjects were eliminated because of procedural error, resulting in a total of 41 subjects.

Materials and stimuli. A list of 592 Eskimo words and their English translations was assembled from various Web sites; the majority of the words were in the Inupiaq dialect (e.g., *ieuieabutaixaq* = nineteen, *uvieiq* = skin). During Exposure 1, the Eskimo word (ranging from 2 to 17 letters in length) was presented in green on a gray background. Words were in capital letters and measured 1 cm × 1.5–9.5 cm. The English translation was presented in red 1.5 cm below the Eskimo word. During testing, the Eskimo word was similarly presented, but the English word was replaced by a cursor so that the subjects could type in their answers.

Design. The experiment had one independent variable: lag condition. Word pairs were randomly assigned (individually for each subject) to one of six lag conditions (1, 2, 4, 8, 16, or 32 intervening items). This determined the desired number of intervening trials between Exposure 2 and Exposure 3 (Test 1 and Test 2). Day 2 performance was analyzed across lag and was conditional on Test 1 performance.

Procedure. Each subject participated in two sessions on consecutive days. Day 1 was a training session. The events of the Day 1 were scheduled at the beginning of the session, as follows. Before beginning the scheduling process, the computer randomized the entire list of word pairs. The study and test trials were then scheduled by iteratively applying the following algorithm:

1. The next word pair was taken from the items remaining on the randomized list of word pairs.

2. The word pair was assigned to be presented on a study trial on the first available open position (call that trial *n*).

3. The first test trial of the word pair (Test 1 for that item) was scheduled to occur on either trial $n + 3$ (lag = 2) or, if that was not available, on the next available open position after that.

4. The nominal lag for the word pair was randomly selected without constraint from the set (1, 2, 4, 8, 16, 32).

5. Test 2 was then scheduled for trial $n + 4 +$ nominal lag or the next available open position.

This was repeated until 200 trials had been scheduled. It was logically impossible to present every word at the precise lag desired because of scheduling conflicts: Another event would often need to occur at the same moment. Rather than include large amounts of filler material, diluting experimental power, we simply allowed the actual lag to increase beyond the nominal lag to the degree necessary to resolve scheduling conflicts. Some items appearing near the end received fewer than the full complement of three presentations; these items were simply omitted from the testing on Day 2. (After we ran the experiment, we discovered that a handful of words on the Eskimo word list were inadvertently mapped onto two different English words, meaning that some subjects would have experienced inconsistent feedback with these items; this affected less than 1% of trials and was unconfounded with variables described here.)

Instructions on the computer screen described the procedure to subjects and told them that on tests, they should respond if reasonably confident. On Exposure 1 (study) of each pair, the Eskimo and English words appeared together in the middle of the screen for 8 s. On test trials, the Eskimo word was presented with a text box below it, which cued the subjects to type in the English word if they were fairly confident that they knew the answer (the text box gave no cues for the number of letters to be typed). Subjects were free to take as long as they needed to respond. After the subject responded, the computer played a sound to indicate whether the response was correct or incorrect. In addition to the audio feedback, if the response was incorrect or the subject clicked "Don't Know," then the correct answer was displayed below the Eskimo word for 5 s. After a 1-s pause accompanied by a blank screen, the computer proceeded to the next trial. The Day 1 session lasted for about 45 min.

On Day 2 (the following day), instructions were again provided by the computer, which also randomized the list of Eskimo words and tested the subject once on each word, presenting the Eskimo word and a text input box where the subject was to type in the English word. Subjects could take as long as they needed to respond, but unlike Day 1, no feedback was provided, and subjects were required to respond to every item. To motivate subjects, we told them that the people who recalled the greatest and second-greatest numbers of words would receive a bonus of $25 and $15, respectively.

Results and Discussion

For short Test 1–Test 2 lags, the algorithm often scheduled Test 2 to occur after a longer lag than was desired, so the mean actual lags noticeably exceeded the short nominal lags (for nominal lags of 1, 2, and 4, the mean actual lags were 2.6, 3.0, and 4.3, respectively). Nominal lags of 8, 16, and 32 were scarcely different from the actual lags of 8.2, 16.1, and 32.0, respectively. In light of this, the results were analyzed and graphed by actual lag for the shorter lags (collapsed into bins of 1–3, 4–6, and 7–9 so as to include adequate amounts of data) and by nominal lag for the longer lags of 16 and 32.

Figure 1 shows overall performance on Test 1 of Day 1, Test 2 of Day 1, and the Day 2 test as a function of lag between Test 1 and Test 2. Test 1 performance was unaffected by lag, confirming that the results were not due to uneven distribution of item diffi-
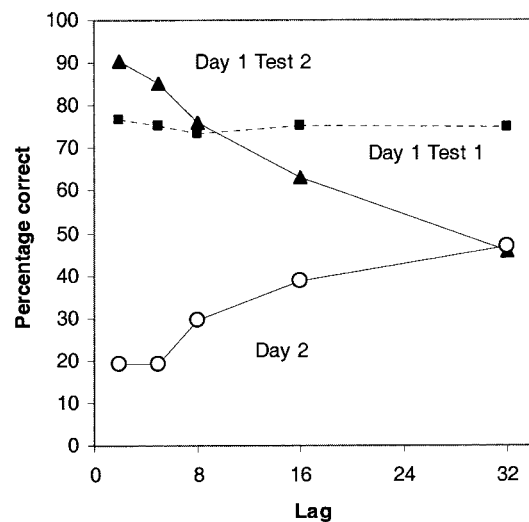

Figure 1. The effect of lag between Test 1 and Test 2 (on Day 1) on accuracy in each of the three tests in Experiment 1. Day 1 Test 2 shows a forgetting within the first day, whereas Day 2 (the final test) shows improved memory with longer spacings.

culty across lags. A comparison of short (≤ 9) and long (≥ 16) lags showed a nonsignificant effect, $F(1, 40) = 0.05$, $p = .83$. As the lag between Test 1 and Test 2 increased, performance in Test 2 fell (a forgetting function). Accuracy was 85% for the short lags and 55% for the long lags, a significant difference, $F(1, 40) = 147.32$, $p < .001$. However, as lag increased, Day 2 performance increased significantly, from 22% for short lags to 43% for longer lags, $F(1, 40) = 89.98$, $p < .001$.

Performance on the final test was conditionalized on accuracy of the subject's response on Test 1 of Day 1. Figure 2 shows performance for items that elicited correct responses in Test 1 versus items that elicited errors (including no response) in Test 1, broken down by lag. An error on Test 1 predicted an error on the final test, possibly due purely to differences in item difficulty. Long lags produced better performance than short lags, whether the subject was correct on Test 1, $F(1, 40) = 65.25$, $p < .001$, or not, $F(1, 40) = 9.60$, $p < .01$.

To guard against statistical anomalies arising from comparing percentage performance from differing sample sizes (e.g., Simpson's paradox; Yule, 1903), we analyzed data in two ways: (a) pooled, with each data point represented equally, and (b) by subject, with each subject contributing a percentage correct score for each bin. The results were essentially identical both qualitatively and quantitatively, indicating that effects described here were not due to statistical anomaly (to save space, only by-subject analyses were reported). In this experiment, we were unable to separate errors of omission from errors of commission because the actual Test 1 response was not stored. Note that the curves in Figure 2, which support the spacing effect, have not obviously peaked at lag $= 32$. Whether the spacing effect strengthens or wanes at longer lags is unclear from this experiment.

Experiment 2

To assess the generality of these findings, a second experiment was conducted. It was similar to Experiment 1 except for three major changes. First, to assess generality to different materials, subjects were taught relatively obscure English vocabulary words similar to those tested on the Graduate Record Exam (GRE). Second, lags of up to 96 intervening words were used (tripling the maximum lag used in Experiment 1). Third, the retest delay was increased from 1 day in Experiment 1 to 1 week in Experiment 2.

Additionally, in Experiment 2, we used a screening procedure to ensure that subjects were not taught words they had already mastered: Whereas in Experiment 1 items were first presented in a study trial, in Experiment 2 the initial exposure was a test. Subjects saw a definition together with the first two letters of a word, and responded by typing in the word. If a subject was able to produce the word, it was not used further with that subject. Last, errors of omission and errors of commission were analyzed separately, to further assess the generality of the findings.

Method

Subjects. Thirty-five students from the University of California, San Diego, participated in two sessions 1 week apart. One subject was eliminated for knowing most of the words, leaving 34 subjects for analysis.

Materials and stimuli. For experiment 2, we used materials identical to those used in Experiment 1, except for two alterations. The range of lags was 1, 4, 16, 32, 64, and 96 intervening items. The stimuli were drawn from a very large pool of relatively infrequent words that were selected from a number of self-study guides for the GRE (e.g., *cygnet, declivity*).[3] Criteria for choosing a word consisted of its relative obscurity and the ability of the experimenters to find a definition which, together with the initial two letters of the word, uniquely picked it out.

Design and procedure. Experiment 2 was procedurally identical to Experiment 1, with three exceptions. As in Experiment 1, there were three Day 1 exposures: the initial exposure (screening test), followed by Test 1 (after a lag of 2) and then Test 2, following the preselected lag. The initial exposure study trials were replaced with a screening test with feedback. If the subject successfully defined the word, that word was eliminated as not novel for that subject. The second session took place 7 days later. We used an improved scheduling algorithm that better succeeded in keeping the actual lag close to the target lag.

Results and Discussion

Data were analyzed as in Experiment 1, including parallel analyses of pooled and per-subject data. As in Experiment 1, results were essentially identical with these different analysis methods. The data shown and analyzed here are by subject.

Actual lags were closer to nominal lags than they were in the preceding experiment (for Nominal Lag 1, the mean lag was 1.27; for Nominal Lag 4, the mean lag was 4.27; for longer nominal lags, the mean differed from the nominal lag by less than 0.20). Therefore, data were graphed and analyzed by nominal lag.

Item difficulty was well spread across time lags, with a mean correct of 56.4% on Test 1 and no difference between short lags of 1 and 4 (56.3%) and longer (≥ 16) lags (56.4%), $F(1, 34) = 0.002$, $p = .96$, *ns*. Screening at initial exposure resulted in excluding about 1.2% words per subject as not novel, and each subject contributed an average of 86.8 scores ($SD = 6.0$; range 73–103). As lag increased, Test 2 performance decreased steadily, from
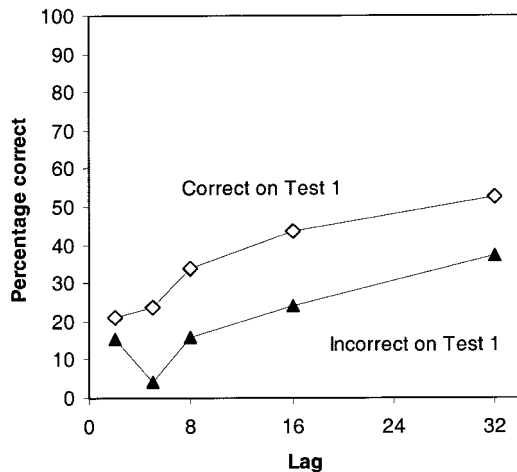

Figure 2. Experiment 1 performance on Day 2 as a function of lag between Test 1 and Test 2 of Day 1 and whether subject responded correctly on Test 1 of Day 1. Results indicate enhanced learning at longer lags (≥ 16), even for items not correctly responded to in Test 1.

[3] The materials are available online at http://www.pashler.com/Definitions.html.

85% correct at lag = 1 to below 30% correct at lag = 96. The curve connecting filled triangles in Figure 3 reflects this forgetting function, with a significant decrease over lag, $F(1, 34) = 285.27$, $p < .01$.

However, as lag was increased, Day 2 (1 week later) performance increased from 20% at lag = 1 to 30% at lag = 96, with a significant difference between short and long (≥ 16) lags, $F(1, 34) = 36.84$, $p < .01$.

Figure 4 shows performance on Day 2 conditionalized on performance on a given item in Test 1 of the initial learning session. Performance was broken down according to whether the subject had responded correctly (55% of trials), failed to respond (18% of trials), or made an error of commission (26% of trials). Errors of commission on Test 1 signaled a better chance of getting the item correct on the final test as compared with errors of omission, $F(1, 34) = 18.51$, $p < .001$.

There was better performance on the final test for long lags (≥ 16) as compared with shorter lags, regardless of success on either of these tests. For correct trials: $F(1, 34) = 22.39$, $p = .001$; for errors of omission: $F(1, 34) = 2.93$, $p < .05$; for errors of commission: $F(1, 34) = 18.83$, $p < .001$.

Although performance is clearly better at the long lags than the short lags, inspection of Figure 4 suggests that optimum learning may occur at intermediate lags between 32 and 96. Given the shallowness of the effects over this range, we were not able to draw any reliable conclusions about the precise optimum or about whether the optimum differed as a function of performance on Test 1. However, our results do make it clear that even after an error on Test 1, it is best to impose a lag that is long enough to produce a great many errors on Test 2 (when subjects made an error of omission on Test 1, after a long Test 1–Test 2 lag they failed to produce the correct answer on Test 2 89% of the time; for errors of commission, they failed 68% of the time).[4]

General Discussion

In two domains with practical relevance (learning foreign language vocabulary and learning obscure English vocabulary), de-
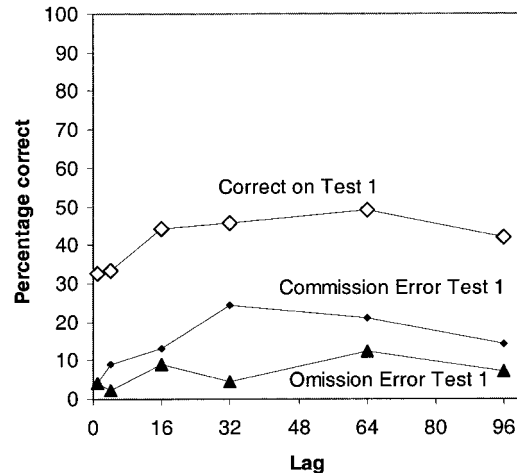
Figure 4. Performance on Graduate Record Exam words (Experiment 2) on the final test after 1-week delay, conditionalized on lag between Test 1 and Test 2 of Day 1 and performance on Test 1 of Day 1. Results indicate enhanced learning at longer lags (≥ 16), even for items not correctly responded to in Test 1, whether errors of omission or errors of commission.

laying a test with feedback dramatically enhanced learning even when the delay was so long as to cause errors to occur on the great majority of trials. The crossover interaction between the effect of lag on same-day versus final test performance (see Figures 1 and 3) provides an excellent illustration of Schmidt and Bjork's (1992) contention that variables that reduce performance during learning may actually enhance learning, as revealed in later tests.

The results provide no evidence of any harmful consequence to making an error per se, at least when feedback follows the error. Errors earlier in learning invariably predict errors later in learning (as they do in the data reported here), but the Guthrian assumption that one learns what one does (even if what one does is to make a mistake) receives little support from these data (even from the errors of commission, which seem most pertinent to Guthrie's proposal). The results suggest, although they do not prove, the opposite possibility: that making errors might actually facilitate the learning process (of course, making one error still predicts subsequent difficulties with that item). It has long been known that errors trigger learning in tasks like *concept learning* (discovering a common rule for classifying diverse instances by scrutinizing examples).[5] Conceivably, the same could be true of associative learning tasks like those examined here, even though there is no common rule to be discovered that can predict the correct response to a not-yet-learned cue.

The results of Experiment 2 provide some hint that the optimum spacing may be less than the maximum lag examined here (96
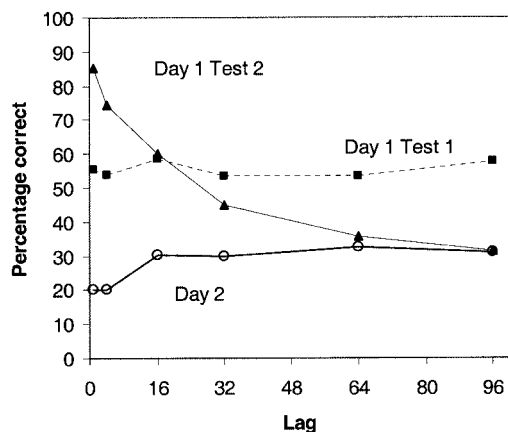

Figure 3. The effect of lag between Test 1 and Test 2 (on Day 1) on each of the three tests of Graduate Record Exam words in Experiment 2. Day 1 Test 2 shows a forgetting function within the first day, whereas Day 2 (the final test, 1 week later) shows improved memory with longer spacings.

---

[4] In both experiments, the same pattern was observed even for trials that elicited errors on both Test 1 and Test 2, although the implications of this are less than straightforward because of item-selection issues.

[5] For example, Trabasso and Bower (1968) summarized that literature as follows: "Opportunities for learning (entering the solution state from the presolution state) occur only in trials in which the subject makes an error, whereas correct response trials provide no opportunity for the subject to exit from the presolution state" (p. 46).

intervening items). Previous writers have suggested that optimum learning takes place when the interstudy interval is some modest proportion of the retention interval (Crowder, 1976), but these writers seem to have had in mind ratios far larger than what were used here (the lag of 96 intervening items corresponds to a duration on the order of 1/1,000 that of the 1-week retention interval). Clearly, the function relating learning to interstudy interval and retention interval is not yet well understood.

Limitations

One obvious limitation of the present results is that, for practical purposes, we are normally interested in achieving learning well above the 50%–60% level attained in this study. Real-world instruction often achieves, at least momentarily, some degree of *overlearning,* that is, more than enough learning to support 100% accurate recall. Overlearning has been shown to retard, or at least delay, subsequent forgetting (Driskell, Willis, & Copper, 1992; Krueger, 1929), but its interactions with interstudy spacing are not known.

A second limitation is that the results were obtained with interleaved trials, in which time is confounded with intervening presentations of other items, possibly generating substantial associative interference. It should not be assumed that the results would necessarily have been the same if the time intervals had been blank or filled with some other unrelated task. This issue could be clarified only with very large amounts of data collection, and from a practical standpoint, the present situation at least approximates many real-world applications (e.g., in computer-aided instruction or in the use of flashcards).

A third limitation, theoretical in nature, is that we cannot draw definitive conclusions about whether making an error is harmful (or helpful) per se. To know what causal impact an error had, uncontaminated by item selection issues, one would need to compare later performance after the subject makes an error on an item with performance on other items for which an error would have been made—but for which no test was even given. Obviously, one has no way of picking out such items. This problem does not, however, stop us from concluding that it is advisable to use delays long enough to provoke a high error rate.

Conclusions

In a vast number of learning situations, ranging from classroom education to computer-aided instruction, from practicing for the GRE to learning a foreign language, one confronts the question of how drills can be optimized. The results described here suggest that using substantial spacing between tests (at least when these are accompanied by feedback) is likely to provide a stronger foundation for optimizing learning than the intuitively appealing idea of arranging conditions to minimize the occurrence of errors.

References

Austin, S. D. M. (1921). A study in logical memory. *American Journal of Psychology, 32,* 370–403.

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science, 4,* 316–321.

Bjork, R. A. (1966). *Learning and short-term retention of paired associ-*

ates in relation to specific sequences of interpresentation intervals (Tech. Rep. No. 106). Stanford, CA: Stanford University Institute for Mathematical Studies in the Social Sciences.

Bjork, R. A. (1988). Retrieval practice. In M. Gruneberg & P. E. Morris (Eds.), *Practical aspects of memory: Current research and issues, Vol. 1: Memory in everyday life* (pp. 396–401). New York: Wiley.

Bregman, A. S., & Wiener, J. R. (1970). Effects of test trials in paired-associate and free-recall learning. *Journal of Verbal Learning & Verbal Behavior, 9,* 689–698.

Butler, D. C., & Peterson, D. E. (1965). Learning during "extinction" with paired associates. *Journal of Verbal Learning & Verbal Behavior, 4,* 103–106.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition, 20,* 632–642.

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

Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology, 14,* 215–235.

Cunningham, D. J., & Anderson, R. C. (1968). Effect of practice time within prompting and confirmation presentation procedures on paired associate learning. *Journal of Verbal Learning and Verbal Behavior, 7,* 613–616.

Dempster, F. N. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. *American Psychologist, 43,* 627–634.

Dempster, F. N. (1996). Distributing and managing the conditions of encoding and practice. In R. Bjork & E. Bjork (Eds.), *Memory* (pp. 317–344). New York: Academic Press.

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

Gagne, R. M., Briggs, L. E., & Wager, W. W. (1992). *Principles of instructional design* (4th Ed.). Fort Worth, TX: Harcourt Brace Jovanovich.

Glenberg, A. M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning and Verbal Behavior, 15,* 1–16.

Glenberg, A. M., & Lehman, T. S. (1980). Spacing repetitions over 1 week. *Memory & Cognition, 8,* 528–538.

Greeno, J. (1964). Paired-associate learning with massed and distributed repetitions of items. *Journal of Experimental Psychology, 67,* 286–295.

Guthrie, E. (1952). *The psychology of learning* (Rev. Ed.). New York: Harper.

Izawa, C. (1970). Optimal potentiating effects and forgetting-prevention effects of tests in paired-associates learning. *Journal of Experimental Psychology, 83,* 340–344.

Karlsson, T. (1989). *A comparison of three prompt-fading methods in computer software training.* Unpublished doctoral dissertation, West Virginia University.

Krueger, W. C. F. (1929). The effect of overlearning on retention. *Journal of Experimental Psychology, 12,* 71–78.

Kulik, C. C., Schwalb, B. J., & Kulik, J. A. (1982). Programmed instruction in secondary education: A meta-analysis of evaluation findings. *Journal of Educational Research, 15,* 133–138.

Kuo, T., & Hirshman, E. (1996). Investigations of the testing effect. *American Journal of Psychology, 109,* 451–464.

Marx, M. H., & Witter, D. W. (1972). Repetition of correct responses and errors as a function of performance reward or information. *Journal of Experimental Psychology, 92,* 53–58.

Pavlov, I. (1927). *Conditioned reflexes.* London: Oxford University Press.

Roediger, H. L., & Payne, D. G. (1982). Hypermnesia: The role of repeated testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 66–72.

Rumelhart, D. E. (1967). *The effects of interpresentation intervals on*

performance in a continuous paired-associate task (Tech. Rep. No. 116). Stanford, CA: Stanford University Institute for Mathematical Studies in the Social Sciences.

Sailer, S. (2002). *Feature: Name game—Inuit or Eskimo?* United Press International wireservice report (Mind and Life Desk). Retrieved June 20, 2002, from http://www.upi.com/view.cfm?StoryID=26062002-104059-8478r

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3,* 207–217.

Skinner, B. F. (1968). *The technology of teaching.* Englewood Cliffs, NJ: Prentice Hall.

Taber, J. I., Glaser, R., & Schaefer, H. H. (1965). *Learning and programmed instruction.* Reading, MA: Addison Wesley.

Trabasso, T., & Bower, G. H. (1968). *Attention in learning: Theory and research.* New York: Wiley.

Vargas, J. S. (1986). Instructional design flaws in computer-assisted instruction. *Phi Delta Kappan, 67,* 738–744.

Young, J. L. (1971). Reinforcement-test intervals in paired-associates learning. *Journal of Mathematical Psychology, 8,* 58–81.

Yule, G. U. (1903). Notes on the theory of association of attributes in statistics. *Biometrika, 2,* 121–134.

Call for Nominations

The Publications and Communications (P&C) Board has opened nominations for the editorships of *Comparative Psychology*, *Experimental and Clinical Psychopharmacology*, *Journal of Abnormal Psychology, Journal of Counseling Psychology*, and *JEP: Human Perception and Performance* for the years 2006–2011. Meredith J. West, PhD, Warren K. Bickel, PhD, Timothy B. Baker, PhD, Jo-Ida C. Hansen, PhD, and David A. Rosenbaum, PhD, respectively, are the incumbent editors.

Candidates should be members of APA and should be available to start receiving manuscripts in early 2005 to prepare for issues published in 2006. Please note that the P&C Board encourages participation by members of underrepresented groups in the publication process and would particularly welcome such nominees. Self-nominations also are encouraged.

Search chairs have been appointed as follows:

- *Comparative Psychology*, Joseph J. Campos, PhD
- *Experimental and Clinical Psychopharmacology*, Linda P. Spear, PhD
- *Journal of Abnormal Psychology*, Mark Appelbaum, PhD, and David C. Funder, PhD
- *Journal of Counseling Psychology*, Susan H. McDaniel, PhD, and William C. Howell, PhD
- *JEP: Human Perception and Performance*, Randi C. Martin, PhD

To nominate candidates, prepare a statement of one page or less in support of each candidate. Address all nominations to the appropriate search committee at the following address:

Karen Sellman, P&C Board Search Liaison
Room 2004
American Psychological Association
750 First Street, NE
Washington, DC 20002-4242

The first review of nominations will begin December 8, 2003. The deadline for accepting nominations is **December 15, 2003.**