H L. , III, J D. K

Washington University in St. Louis

AB AC

In educational settings, tests are usually considered devices of assessment. Students take tests in class to assess what they have learned and take standardized tests like the SAT to assess their knowledge and aptitude. In many circumstances, such as university lecture courses, tests are given infrequently (often just two or three times a semester) and are generally perceived as a bother by faculty and students alike. We believe that the neglect of testing in all levels of education is misguided. To state an obvious point, if students know they will be tested regularly (say, once a week, or even every class period), they will study more and will space their studying throughout the semester rather than concentrating it just before exams (see Bangert-Drowns, Kulik, & Kulik, 1991; Leeming, 2002). However, more important for present purposes, testing has a powerful positive effect on future retention. If students are tested on material and successfully recall or recognize it, they will remember it better in the future than if they had not been tested. This phenomenon, called the testing effect, has been studied sporadically over a long period of time (e.g., Gates, 1917), but is not well known outside cognitive psychology.

took a test on the material or studied it again before taking a final retention test 5 min, 2 days, or 1 week later. In Experiment 2, students studied a passage once and took three tests, studied

twice recalled more than subjects who had studied once and taken a recall test. However, this pattern of results was reversed on the delayed tests 2 days and 1 week later. On these tests of long-term retention, subjects who had taken an initial test recalled more than subjects who had only studied the passages.

The results were submitted to a 2 × 3 analysis of variance (ANOVA), with learning condition (restudying or testing) and retention interval (5 min, 2 days, or 1 week) as independent variables. This analysis revealed a main effect of testing versus restudying, $F(1, 117) = 36.39$, $\eta_p^2 = .24$, which indicated that, overall, initial testing produced better final recall than additional studying. Also, the analysis revealed a main effect of retention interval, $F(2, 117) = 50.34$, $\eta_p^2 = .46$, which indicated that forgetting occurred as the retention interval grew longer. However, these main effects were qualified by a significant Learning Condition × Retention Interval interaction, $F(2, 117) = 32.10$, $\eta_p^2 = .35$, indicating that restudying produced better performance on the 5-min test, but testing produced better performance on the 2-day and 1-week tests.

Post hoc analyses confirmed that on the 5-min retention tests, restudying produced better recall than testing (81% vs. 75%), $t(39) = 3.22$, $d = 0.52$. However, the opposite pattern of results was observed on the delayed retention tests. After 2 days, the initially tested group recalled more than the additional-study group (68% vs. 54%), $t(39) = 6.97$, $d = 0.95$. The benefits of initial testing were also observed after 1 week: The tested group recalled 56% of the material, whereas the restudy group recalled only 42%, $t(39) = 6.41$, $d = 0.83$. Figure 1 depicts another interesting finding: The initially tested group recalled as much on the 1-week retention test as the additional-study group did after only 2 days (the initially tested group actually recalled slightly more). This surprising result indicates that taking an initial recall test prevented forgetting of information for an additional 5 days relative to repeated study.

periods, subjects were given a blank sheet and were asked to recall as much of the material from the passage as they could remember, without concern for exact wording or correct order. Each test lasted 10 min, and subjects were instructed to draw a line on their test sheets to mark their place after each 1-min interval. Subjects solved multiplication problems for 2 min between periods and for 5 min after the final period in Phase 1.

At the end of Phase 1, subjects were given a questionnaire asking them to answer three questions using a 7-point scale. They indicated how interesting they thought the passage was (1 = very boring, 7 = very interesting), how readable they thought it was (1 = very easy to read, 7 = very difficult to read), and how well they thought they would remember the passage in 1 week (1 = not very well, 7 = very well). After completing the questionnaire, subjects in the 5-min retention-interval condition took the final recall test, and subjects in the 1-week condition were excused, returning for the final test 1 week later. The final recall test (Phase 2) was identical to the initial recall tests.

Readings of the Passage

The mean number of times subjects were able to read through the passage during each study period is presented in Table 1. No differences in these reading scores were observed for the two passages or for the 5-min and 1-week retention-interval groups. Across all conditions, subjects were able to read the entire passage approximately 3.5 times during a 5-min study period. The number of times subjects in the SSSS and SSST conditions read the passage increased slightly across consecutive study periods, $F(3, 177) = 1.62$, $\eta_p^2 = .03$, and $F(2, 118) = 4.99$, $\eta_p^2 = .08$, respectively. The reading scores in Table 1 simply illustrate that subjects read the passage many more times in the SSSS ($M = 14.2$) and SSST ($M = 10.3$) conditions than in the STTT ($M = 3.4$) condition.

Initial Tests

Subjects in the STTT condition recalled 20.9, 21.2, and 21.1 idea units on each of the three initial recall tests, respectively, or about 70% of the passage in each case. No differences on the initial tests were observed for the two passages or for the 5-min and 1-week retention-interval groups. Measures of cumulative recall indicated that asymptotic levels of recall had been reached by the end of each test period. Subjects in the SSST condition recalled 23.1 idea units (77% of the passage) on their initial recall test. This was reliably greater recall than on the third test in the STTT condition, $t(118) = 3.17$, $d = 0.58$.

TABLE 1

Questionnaire

The mean ratings on the questionnaire given at the end of Phase 1 are displayed in Table 2. No differences in the questionnaire ratings were observed for the two passages or for the 5-min and 1-week retention-interval groups. Subjects in the SSSS condition rated the passage as less interesting than subjects in the SSST or STTT condition, $F(2, 177) = 3.88$, $\eta^2 = .04$, perhaps because of increased boredom with repeated readings. More interestingly, subjects in the SSSS condition were more confident that they would remember the passage in 1 week than were subjects in the SSST or STTT condition, $F(2, 177) = 6.09$, $\eta^2 = .06$. Post hoc analyses revealed that subjects in the SSSS condition predicted that they would remember the passage better than subjects in the SSST condition, $t(118) = 2.95$, $d = 0.54$, and subjects in the STTT condition, $t(118) = 3.35$, $d = 0.61$, but the latter two groups did not differ significantly in their predictions. The three groups did not differ in how they rated the readability of the passages ($F < 1$).

Final Tests

The critical data are the mean proportions of idea units recalled on the final tests 5 min or 1 week later, displayed in Figure 2. The pattern of final test scores replicates the pattern of results found in Experiment 1. On the 5-min test, recall was correlated with repeated studying: The SSSS group recalled more than the SSST group (83% vs. 78%), who in turn recalled more than the STTT group (71%). However, on the 1-week test, recall was correlated with the number of tests given earlier: The STTT group recalled more than the SSST group (61% vs. 56%), who in turn recalled more than the SSSS group (40%).

The results in Figure 2 were submitted to a 2×3 ANOVA, with retention interval (5 min or 1 week) and learning condition

tests, whereas spaced presentation leads to better performance on delayed tests (Balota et al., 1989; Peterson, Wampler, Kirkpatrick, & Saltzman, 1963). That is, in both cases, massed study leads to a short-term benefit, but the other manipulation (testing or spaced studying) has a greater effect on long-term retention. Both outcomes may reflect the role of desirable difficulties in promoting long-term retention (Bjork, 1994), as discussed later. This outcome on the immediate tests in the present experiments reveals just how powerful the testing effect is: Despite the benefits of repeated study shortly after learning, repeated testing produces strong positive effects on a delayed test.

Several overlapping theoretical approaches are useful in understanding our results. The findings are consistent with theories of transfer-appropriate processing that emphasize the compatibility between the operations engaged in during learning and testing phases (Morris, Bransford, & Franks, 1977; Roediger, 1990). The ability to remember a prose passage on a free-recall test a week after learning it is enhanced by practicing exactly this skill during learning. Practicing the skills during learning that are needed during retrieval generally enhances retention on both explicit and implicit memory tests (Roediger, Gallo, & Geraci, 2002). Although restudying the passages exposed students to the entire set of information, testing permitted practice of the skill required on future tests and hence enhanced performance after a delay.

McDaniel and his colleagues (McDaniel & Fisher, 1991; McDaniel, Kowitz, & Dunay, 1989; McDaniel & Masson, 1985) have argued that testing enhances learning by producing elaboration of existing memory traces and their cue-target relationships, and Bjork (1975, 1988) has suggested that testing operates by multiplying the number of ''retrieval routes'' to stored events. Bjork (1994, 1999) has also emphasized the need to introduce desirable difficulties into training and educational settings. Many study conditions and strategies that produce rapid learning and short-term benefits lead to poor long-term performance. Our results show that testing versus studying is another case in point: Testing clearly introduced a desirable difficulty during learning.

Relative to testing, repeated studying inflated students' confidence in their ability to remember the passages in the future, even though repeated-study subjects actually showed much poorer retention on delayed tests. Repeated studying is a strategy that students frequently report using and is often recommended to students by teachers (see Rawson & Kintsch, 2005, for discussion). Students may prefer repeated studying because it produces short-term benefits, and students often use ineffective learning strategies because they base their predictions of future performance on what produces rapid short-term gains. Although students in the repeated-study condition predicted they would perform very well a week later (relative to those in the other conditions), they actually performed the worst.

Free-recall testing even without feedback had large positive effects on retention in our experiments. e6.3(ed)-17g.3(ca9TJ0-1.22ed)-theytermtud. t-435.3lhwl1ing379.9(our)-3ted9 is

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. Memory & Cognition, 20, 633–642.

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

Gates, A.I. (1917). Recitation as a factor in memorizing. Archives of Psychology, 6(40).

Glover, J.A. (1989). The ''testing'' phenomenon: Not gone but nearly forgotten. Journal of Educational Psychology, 81, 392–399.

Hogan, R.M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. Journal of Verbal Learning and Verbal Behavior, 10, 562–567.

Izawa, C. (1967). Function of test trials in paired-associate learning. Journal of Experimental Psychology, 75, 194–209.

Koriat, A., Bjork, R.A., Sheffer, L., & Bar, S.K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. Journal of Experimental Psychology: General, 133, 643–656.

Landauer, T.K., & Bjork, R.A. (1978). Optimum rehearsal patterns and name learning. In M.M. Gruneberg, P.E. Morris, & R.N. Sykes