tapraid5/zfv-xlm/zfv-xlm/zfv00608/zfv2229d08g | sangreyj | S=12 | 12/2/08 | 17:09 | Art: 2008-0940

Journal of Experimental Psychology:
Learning, Memory, and Cognition
2008, Vol. 34, No. 6, 1392–1399

0278-7393/08/$12.00 DOI: 10.1037/a0013082

Testing During Study Insulates Against the Buildup of Proactive Interference

Karl K. Szpunar, Kathleen B. McDermott, and Henry L. Roediger III
Washington University in St. Louis

Recent interest in the benefits of retrieval practice on long-term retention—the testing effect—has spawned a considerable amount of research toward understanding the underlying nature of this ubiquitous memory phenomenon. Taking a test may benefit retention through both direct means (engaging appropriate retrieval processes) and indirect means (fostering directed study). Here the authors report 4 experiments demonstrating a novel benefit of testing. Extended study sessions cause a buildup of proactive interference, but interpolating tests during the study sequence insulates against this negative influence. These findings highlight a unique benefit of testing and have important implications for study strategies.

Keywords: testing effect, proactive interference, list discrimination, cue overload, source monitoring

Testing one's memory enhances long-term retention (Glover, 1989). In a recent review of the literature, Roediger and Karpicke (2006a) identified two means by which testing functions as an effective study tool. First, the act of taking a test itself appears to *directly* benefit retention. For instance, after studying a set of materials (e.g., a list of words, lecture notes), one is more likely to remember those materials at a later time if the study session is followed by an immediate memory test (e.g., Spitzer, 1939). In fact, the long-term benefit of testing outweighs that of additional study time (Roediger & Karpicke, 2006b). According to current theories of testing, an initial test will directly benefit retention to the extent that it engages those effortful retrieval processes that will likely be called upon at the time of a later test (Roediger & Karpicke, 2006a). Secondly, given the opportunity for repeated study, learners may use a test as feedback in order to make better use of subsequent study sessions (Thompson, Wenger, & Bartling, 1978). That is, a test may *indirectly* enhance study strategies by allowing the learner to focus additional study time on yet-to-be-acquired (i.e., not recalled or erroneously recalled) material.

In the present report, we outline an additional function of testing that has received little attention to this point but has potentially important implications for study strategies. In both the laboratory and the classroom, learners are frequently required to memorize consecutive sets of verbal materials (e.g., lists of words, lecture notes) in anticipation of a later test (e.g., final free recall, class exam). Moreover, these extended study sessions typically take place in a single study session (Indig, 2005; Leeming, 2002; Michael, 1991). It is well recognized, however, that the successive acquisition of verbal materials is a sufficient condition for buildup of proactive interference (PI; Postman & Keppel, 1977). That is, successful retention of new information is inversely related to the number of prior-learning episodes (Underwood, 1957). Hence, any factor that might moderate the negative influence of prior learning on current learning should be of considerable interest. Several studies have reported data hinting that administering intermittent tests during the course of a multilist learning experiment may insulate learners against the buildup of PI. We review the relevant evidence below and provide new experiments to document the claim more fully.

Darley and Murdock (1971) were the first to report data suggesting that learning a new set of materials may be influenced by whether or not a preceding set of materials had been tested. In their experiment, participants learned 10 consecutive lists of words in anticipation of a final cumulative test. Half of these lists were followed by an immediate recall test, whereas the other half were followed by a distracter task. The only restriction was that no more than 3 consecutively tested or nontested lists followed one another. Darley and Murdock observed that participants sometimes produced prior list intrusions when attempting to recall a list they had just studied (Zaromb et al., 2006). Of particular interest, the authors noted that 81% of prior-list intrusions were words from previously nontested lists, indicating that previously recalled lists interfered less with the learning of a new list than did nontested lists.

Several years later, Tulving and Watkins (1974) reported a peculiar finding using the classic retroactive interference (RI) paradigm (Barnes & Underwood, 1959). In the typical experiment, participants begin by studying one paired-associate list (A–B), taking a test on that list (A–?), and then repeating this procedure with a second, related paired-associate list (A–C). A final test is then administered during which participants are asked to recall each of the response terms (B and C) that had been paired with each stimulus term (A). Generally, participants are better able to remember the more recent response terms (i.e., C > B; Keppel, 1968). Tulving and Watkins discovered that they could reverse this pattern by removing the initial tests that followed the presentation

Karl K. Szpunar, Kathleen B. McDermott, and Henry L. Roediger III, Department of Psychology, Washington University in St. Louis.

This research was supported by a grant from the James S. McConnell Foundation.

Correspondence concerning this article should be addressed to Karl K. Szpunar, Department of Psychology, Washington University in St. Louis, Campus Box 1125, St. Louis, MO 63130. E-mail: karl.szpunar.wustl.edu

of each list (see also Malis, 1970). In the absence of initial testing, participants were better able to remember the response terms they had first encountered (i.e., B > C). Tulving and Watkins further demonstrated that this *priority effect* was the result of impaired second-list acquisition. Specifically, second-list learning, as inferred by an immediate test, was impaired if the first list had not been tested (Akers & Lyons, 1979; Allen & Arbak, 1976; Robbins & Irvin, 1976).

In a recent experiment from our own laboratory (Szpunar, McDermott, & Roediger, 2007), participants studied five 18-word lists in preparation for a final cumulative test. One group studied the five lists with short breaks but no tests immediately after study; a second group was tested after each of the five lists. Relative to the nontested control group, providing initial tests substantially improved long-term retention. This was particularly apparent for materials learned later in the study sequence. That is, participants in the tested group were better able to recall lists they had most recently learned (for a thorough examination of this pattern of results, see Tulving & Psotka, 1971; Tulving & Thornton, 1959). However, nontested participants showed the opposite pattern (i.e., a list primacy effect). Nontested participants learned the first list relatively well, but final recall of lists learned later in the study sequence was markedly impaired in the absence of testing during study. Although we speculated that the presence of testing during study might have determined how well these later study lists were learned, there was no measure of initial learning to support our claim.

The aim of the four experiments reported here was to examine directly the assertion that testing throughout the study phase serves to insulate against the influence of PI. To this end, we developed a novel paradigm based on the experiment reported by Szpunar et al. (2007). In each experiment, participants studied five word lists in anticipation of a final cumulative test. The expectation of a final test ensured the continued processing of materials across the study sequence. Experiments 1A (interrelated lists) and 1B (unrelated lists) examined initial and final recall performance for the last list in the study sequence (i.e., List 5), comparing participants who had or had not been initially tested for the previous study lists (i.e., Lists 1–4). If a test does insulate against the influence of PI, then we should observe two patterns of results. The group of participants who are tested after each list should initially recall each list equally well; further, recall of the last study list (List 5) should be impaired for the previously nontested participants, relative to the participants who had taken four preceding tests. Indeed, if PI does exert a negative influence on learning, then this latter pattern of results should be observed in both initial and final recall. Experiment 2 investigated whether PI increases over time, in the absence of testing. If PI does increase in the absence of testing, then list recall should be inversely related to the number of previously nontested lists encountered in this paradigm. Experiment 3 examined whether the act of retrieval, and not the reexposure to material that accompanies a test, was responsible for insulating against the influence of PI.

Experiments 1A and 1B

With the exception of materials, all procedures were the same for Experiments 1A and 1B; therefore, the methods for these experiments are presented together.

Method

Participants. Twenty-four Washington University undergraduates participated in each experiment for partial fulfillment of a course requirement. The participants were randomly assigned to one of two experimental conditions, resulting in 12 participants per condition.

Materials. In Experiment 1A, five interrelated lists of 18 words were constructed. Each list was composed of three words from each of six semantic categories taken from the Van Overschelde, Rawson, and Dunlosky (2004) category norms. The six categories common to each list were building parts, earth formations, animals, fruits, human body parts, and weather phenomena. The first 5 exemplars from each category were omitted to reduce the likelihood of participants correctly guessing words during testing. The next 15 exemplars (6–20) in each category were divided into five groups of 3 words each. These category triads were then assembled to form the five word lists. It was expected that semantically related materials would ensure buildup of PI across word lists (Wickens, 1970, 1972). However, to ensure that any significant findings were not specific to our choice of materials, Experiment 1B used 90 unrelated words (concrete nouns) of medium frequency drawn from the Kučera and Francis (1967) norms. These words were broken down into five lists of 18 words each. Experimental materials were presented using E-Prime software.

Procedure. Participants were initially informed that the experiment was designed to test their memory and mathematical abilities. The five 18-word lists were visually presented in the center of a computer monitor. The study words were presented at 2 s per word (500-ms interstimulus interval, 2,500-ms stimulus onset asynchrony), and participants were instructed to pay close attention to each word for an upcoming test. List and word order were randomized across participants.

Participants were told to expect to complete two tasks following the presentation of each study list. First, they solved math problems for 1 min, to minimize primary memory effects (Glanzer & Cunitz, 1966). Participants then either completed a 2nd min of math or were given 1 min to recall (in any order) as many words as possible from the list they had just studied by typing in responses. Participants were told that the number of lists for which they would receive an initial free-recall test would be determined randomly and that they should study each list as if they expected a test, in order to maximize performance. In fact, half of the participants were tested after each list, whereas the other half was tested for only the final study list (i.e., List 5). The critical comparison between the two groups was List 5 recall performance. Finally, to encourage participants to continue processing materials across the study sequence, we told them to expect a final cumulative free-recall test that would be administered approximately 30 min after completion of the initial five study (test) segments. The retention interval included participation in an unrelated verbal exercise. During the final test, participants were allotted 8 min to recall as many words as they could from all five lists of words they had studied. Responses were handwritten. It was stressed that the participants were to use all 8 min efficiently in their attempt to recall study materials. The experiment was completed in approximately 1 hr by all participants, at which point they were thanked for their participation and fully debriefed.

tapraid5/zfv-xlm/zfv-xlm/zfv00608/zfv2229d08g | sangreyj | S=12 | 12/2/08 | 17:09 | Art: 2008-0940

Results for Experiment 1A

Both correct recall and intrusion rates (Darley & Murdock, 1971; Underwood, 1945) are considered. To facilitate comparisons between these measures, all data are presented in terms of raw scores. Alpha level was set at .05. Cohen's *d* indicates effect size for *t* tests.

Initial tests. Participants who had been tested following each list recalled approximately 50% of each study list ($M = 9.63$, range: 9.17–9.92 words) and produced few prior-list intrusions in recalling study Lists 2–5 ($M = 0.40$, range: 0.17–0.58 intrusions; $Fs < 1$). Previous learning had no influence on later learning if tests were given after each list (Szpunar et al., 2007; Tulving & F1 Watkins, 1974). However, as can be seen in the left side of Figure 1, the absence of testing, for Lists 1–4, was associated with impaired performance on List 5. Participants who had not been tested following Lists 1–4 recalled significantly fewer List 5 words ($M = 4.40$ vs. 9.67), $t(22) = 3.95$, $d = 1.69$, and made more prior-list intrusions ($M = 3.75$ vs. 0.42), $t(22) = 3.07$, $d = 1.53$.

Final test. Participants who had been initially tested following each list recalled reliably more words on the final cumulative free-recall test than participants who had not been initially tested on Lists 1–4 ($M = 46.42$ and 24.75 for those tested and not tested, respectively), $t(22) = 5.71$, $d = 2.43$. Of particular interest was the number of List 5 words these groups of participants recalled on the final test. To this end, final recall protocols were examined for the presence of List 5 words on a subject-by-subject basis. Mean List 5 recall is presented in the right side of Figure 1. As with initial recall performance, participants who had not been tested following Lists 1–4 recalled significantly fewer List 5 words than those who had been tested ($M = 4.67$ vs. 10.00), $t(22) = 5.09$, $d = 2.17$, on the final cumulative test.

Results for Experiment 1B

Initial tests. Participants who had been tested following each list recalled approximately 40% of each study list ($M = 7.40$, range: 6.80–8.20 words) and produced few prior-list intrusions in recalling Lists 2–5 ($M = 0.25$, range: 0.20–0.30 intrusions; $Fs < 1$). As in Experiment 1A, previous learning had no influence on later learning in the presence of testing. However, as can be seen in the left side of Figure 2, the absence of previous testing was F2 associated with impaired performance on a later learning episode. Participants who had not been tested following Lists 1–4 recalled significantly fewer List 5 words ($M = 3.50$ vs. 7.00), $t(22) = 3.00$, $d = 1.35$, and made more prior-list intrusions ($M = 3.70$ vs. 0.30), $t(22) = 2.89$, $d = 1.57$.

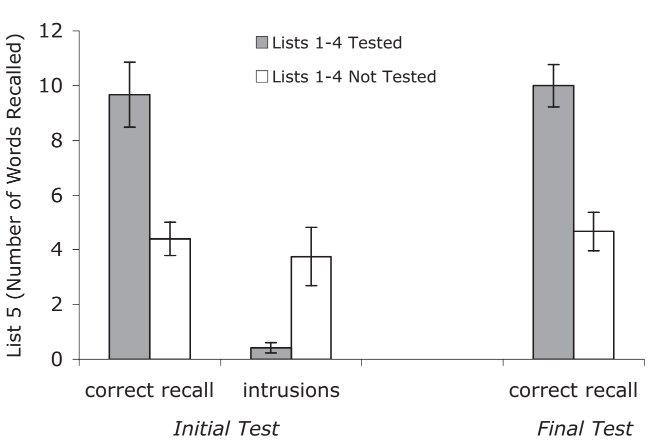

Figure 1. Mean initial and final List 5 recall performance in Experiment 1A as a function of condition. Error bars display standard errors of the mean.

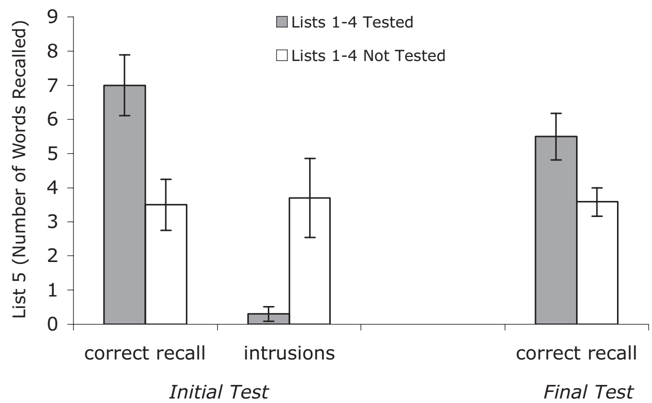

Figure 2. Mean initial and final List 5 recall performance in Experiment 1B as a function of condition. Error bars display standard errors of the mean.

Final test. Participants who had been initially tested following each list recalled reliably more words on the final cumulative test than participants who had not been initially tested on Lists 1–4 ($M = 23.50$ and 16.92 for those tested and not tested, respectively), $t(22) = 2.53$, $d = 1.08$. Again, particular interest was directed toward the number of List 5 words these groups of participants recalled on the final cumulative test. Mean List 5 recall is presented in the right side of Figure 2. As with initial recall performance, participants who had not been tested following Lists 1–4 recalled significantly fewer List 5 words than those who had been tested ($M = 3.58$ vs. 5.50), $t(22) = 2.40$, $d = 1.02$, on the final cumulative test.

These results replicate previous findings demonstrating that nontested list words are more likely to occur as prior-list intrusions (Darley & Murdock, 1971). More important, the data indicate that testing previously learned materials insulates against the negative influence that those materials can have on later learning (Tulving & Watkins, 1974). These patterns emerged with both interrelated and unrelated stimulus materials. Experiments 2 and 3 used interrelated word lists.

Experiment 2

The purpose of the second experiment was to track performance as a function of the amount of PI. Five groups of participants were tested. As before, one group of participants received a test after each list. The four remaining groups of participants were tested for only one list in the study sequence—List 2, List 3, List 4, or List 5. As such, these latter groups differed in the number of previously nontested lists encountered prior to the test—one, two, three, or four, respectively. Figure 3 illustrates the design of this F3 experiment.

If PI accrues strength in the absence of immediate testing, two patterns of results should emerge. First, participants tested after

Group	List 1		List 2		List 3		List 4		List 5	
Always Tested	S	**T**	S	**T**	S	**T**	S	**T**	S	**T**
1 Previously Non-Tested List	S		S	**T**	S		S		S	
2 Previously Non-Tested Lists	S		S		S	**T**	S		S	
3 Previously Non-Tested Lists	S		S		S		S	**T**	S	
4 Previously Non-Tested Lists	S		S		S		S		S	**T**

Figure 3. A schematic of the design of Experiment 2. S = study; T = test.

each list should reveal no difference across lists during initial recall performance. Second, recall performance should be inversely related to the number of previously nontested lists encountered. To address this prediction, we compared list recall for each of the four groups receiving one test with the corresponding list in the group that was tested after each list. That is, the influence of one previously nontested list was determined by comparing List 2 recall for the group that had studied one previously nontested list before their test with List 2 recall for the group that had been previously tested on that list (see Figure 3). The influence of two previously nontested lists was determined by comparing List 3 recall for the group that had studied two previously nontested lists before their test with List 3 recall for the group that had been previously tested on those lists, and so on for three and four previously nontested lists.

Method

Participants. Sixty undergraduate students participated. The participants were tested in small groups that were assigned at random to one of five experimental conditions, resulting in 12 participants per condition.

Materials. The materials were identical to those used in Experiment 1A.

Procedure. All procedures and instructions presented to participants were identical to those of Experiments 1A and 1B.

Results and Discussion

Initial tests. Participants who had been tested following each list recalled approximately 50% of each study list (M = 9.63, range: 9.50–9.70 words) and produced few prior-list intrusions in recalling Lists 2–5 (M = 0.53, range: 0.30–0.70 intrusions; Fs < 1). As before, previous learning episodes had no influence on later learning in the presence of testing.

F4 As can be seen in Figure 4A, as the number of previously nontested lists increased, correct recall of a newly learned set of materials decreased. Relative to the previously tested group, one previously nontested list had a nonsignificant influence on List 2 recall ($M_{diff} = -0.5$ words), although it was in the expected direction, $t(22) < 1$. However, this difference became systematically larger after participants had encountered two ($M_{diff} = -3.40$), three ($M_{diff} = -3.60$), and four ($M_{diff} = -4.40$) previously nontested study lists, $t(22) = 2.46$, $d = 1.10$; $t(22) = 2.54$, $d = 1.16$; and $t(22) = 4.98$, $d = 2.22$, respectively.

As can be seen in Figure 4B, the opposite pattern was obtained with intrusion rates. That is, as the number of previously nontested lists increased, so did the number of prior-list intrusions. Relative to the previously tested group, one previously nontested list had a nonsignificant influence on intrusion rate ($M_{diff} = 1.10$ intrusions), although it was in the expected direction, $t(18) = 1.60$, $p > .10$. As with correct recall, this difference became systematically larger after participants had encountered two ($M_{diff} = 2.20$), three ($M_{diff} = 2.60$), and four ($M_{diff} = 3.00$) previously nontested study

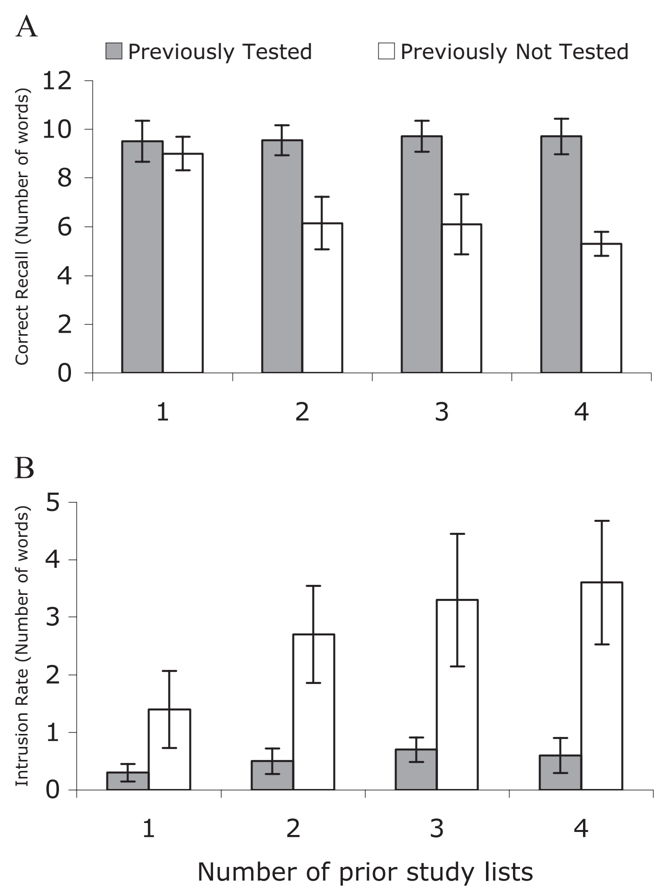

Figure 4. Mean recall performance (correct recall in 4A and intrusion rate in 4B) in Experiment 2 as a function of number of prior study lists and condition. Error bars display standard errors of the mean.

lists, $t(18) = 2.21$, $d = 1.29$; $t(18) = 2.52$, $d = 1.73$; and $t(18) = 2.68$, $d = 2.46$, respectively.

Final test. Participants who had been initially tested following each list exhibited variability in the number of words they recalled from each list on the final cumulative test, $F(4, 44) = 3.93$, $MSE = .005$, $\eta_p^2 = .263$. In a replication of the findings in Szpunar et al. (2007), these participants recalled the most recently studied list best ($M = 9.90$ words) and the most remotely studied list worst ($M = 6.97$ words), $t(11) = 2.48$, $d = 1.17$. That is, there was a list recency effect, which was further corroborated by a significant linear trend in the data, $F(1, 11) = 5.90$, $MSE = .165$, $\eta_p^2 = .349$.

As discussed briefly in the introduction, a number of studies have documented the finding that previously tested word lists exert a retroactive influence at the time of a final cumulative test. Although a thorough discussion of this phenomenon is beyond the scope of the present article (see Tulving & Psotka, 1971; Tulving & Thornton, 1959), it is relevant in considering interpretations of the final recall data of this experiment. Specifically, comparisons of the previously tested group with respective previously nontested groups (as was done with the initial recall data) would likely bias our data in the direction of our predicted results—reduced recall performance with increasing amounts of previously nontested information. That is, comparisons of lists early in the study sequence would reflect little PI for the previously nontested group and a considerable amount of RI for the previously tested group. Comparisons of lists late in the study sequence would reflect considerable amounts of PI for the previously nontested group and relatively little RI for the previously tested group.

Nonetheless, because List 5 was not subject to RI, we were still able to make a nonbiased assessment of the influence of PI on long-term retention by comparing the number of List 5 words recalled on the final test by participants who had and had not been tested following Lists 1–4. As was the case in Experiments 1A and 1B, participants who had not been tested following Lists 1–4 recalled significantly fewer List 5 words than those who had been tested ($M = 6.55$ vs. 9.90), $t(22) = 3.29$, $d = 1.40$.

The results of this experiment are notable in three respects. First, they replicate and extend the findings of Experiments 1A and 1B. Second, the data reveal a striking effect in learning successive word lists: As the number of previously nontested lists increases, the number of words correctly recalled from a new list decreases and the number of prior-list intrusions increases. Finally, in order to produce clear interpretations of the role of PI, we had to ensure that the influence of RI did not produce confounds in the data. Although such considerations were necessary for present purposes, future studies might examine whether PI and RI interact in interesting ways to affect long-term retention. The multilist learning paradigm presented here seems like it might be well suited to address this question.

Experiment 3

It is important to note that the results reported thus far might not be related to testing per se. Specifically, prior testing may somehow benefit later learning from the reexposure to study materials that accompanies testing and not from the act of retrieval itself (Roediger & Karpicke, 2006b). The purpose of Experiment 3 was to address this possibility directly. Three groups of participants were tested. As in Experiments 1A and 1B, one group of participants was tested after each list and one group was tested for only List 5. In addition, a third group of participants was also tested for only List 5 but was allowed to restudy Lists 1–4. If our previous results were due to reexposure, then participants who restudy Lists 1–4 should learn List 5 as well as participants who are tested following Lists 1–4. Indeed, participants who restudy might be expected to outperform participants who are tested, given that they would be reexposed to all study materials, whereas the previously tested participants would be reexposed to only those materials that they could recall (Roediger & Karpicke, 2006a, 2006b). Alternatively, if our previous findings are due to the act of retrieval, then the previously tested participants should outperform both participants who had not been tested following Lists 1–4 and the restudy group.

Method

Participants. Thirty-six undergraduate students participated. The participants were tested in small groups that were assigned at random to one of three experimental conditions, resulting in 12 participants per condition.

Materials. The materials were identical to those used in Experiment 1A.

Procedure. All procedures are identical to those of Experiments 1A and 1B, with the following exception. Presentation time for study words was increased from 2 to 2.5 s per word (500-ms interstimulus interval, 3,000-ms stimulus onset asynchrony). As a result, extra study time (54 s) was approximately equated with testing and extra math episodes (both 1 min).

Instructions were identical to those in Experiments 1A and 1B, with the following exception. To equate study strategies, we informed all participants that they would receive a free recall test, an additional study session (restudy list in a new random order), or additional math after studying each list and completing the first set of math problems that followed. As before, participants were further told that the number of lists for which they would receive an initial test (or additional study) would be determined randomly and that they should study each list as if they expected a test, in order to maximize performance.

Results and Discussion

Initial tests. Mean List 5 recall is presented in the left side of Figure 5. One-way analyses of variance (ANOVAs) revealed significant main effects of condition for both correct recall, $F(2, 33) = 9.91$, $MSE = .036$, partial $\eta_p^2 = .375$, and intrusion rate, $F(2, 33) = 5.70$, $MSE = .033$, $\eta_p^2 = .257$. Between-groups comparisons revealed that participants who had been tested following Lists 1–4 correctly recalled more List 5 words ($M = 9.67$) than both the previously nontested participants ($M = 4.75$) and the participants who had restudied the earlier lists ($M = 3.92$), $t(22) = 3.51$, $d = 1.47$, and $t(22) = 3.74$, $d = 1.53$, respectively. Conversely, participants who had been tested following Lists 1–4 produced fewer prior-list intrusions ($M = 0.58$) than both the previously nontested participants ($M = 4.08$) and the participants who had restudied those lists ($M = 4.75$), $t(22) = 2.63$, $d = 1.31$, and $t(22) = 4.35$, $d = 2.08$, respectively. The previously nontested and restudy groups did not differ from one another in either respect. F5

Final test. A one-way ANOVA revealed a significant main effect of condition in final cumulative recall, $F(2, 33) = 5.44$,

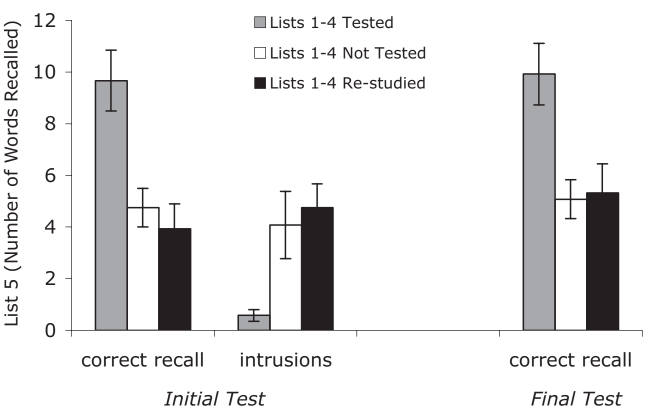

Figure 5. Mean initial and final List 5 recall performance in Experiment 3 as a function of condition. Error bars display standard errors of the mean.

MSE = .025, η_p^2 = .247. Of particular interest was the number of List 5 words these groups of participants recalled on the final cumulative test. Mean List 5 recall is presented in the right side of Figure 5. As with initial recall performance, a one-way ANOVA revealed a significant main effect of condition for correct recall, $F(2, 33) = 6.84$, $MSE = .040$, $\eta_p^2 = .293$. Between-groups comparisons revealed that participants who had been initially tested following each list recalled more List 5 words on the final cumulative test (M = 9.92) than participants who had not been tested following Lists 1–4 (M = 5.08), $t(22) = 3.42$, d = 1.46, and participants who had restudied Lists 1–4 (M = 5.33), $t(22) = 2.80$, d = 1.19. There were no reliable differences between the nontested and restudy groups on recall of List 5 words. The present results indicate that our findings can be attributed to the act of retrieval.

General Discussion

In four experiments, we have highlighted a unique benefit of testing when administered during the course of study. Our findings are notable in at least two respects. First, testing previously learned materials insulates against the negative influence that those materials can have on learning a new set of materials (cf. Sahakyan, Delaney, & Kelley, 2004; Tulving & Watkins, 1974). This finding appeared in all four experiments. Second, this benefit of testing can be attributed to the act of retrieval and not to the reexposure of materials that accompanies testing (Experiment 3). In a recent review, Roediger and Karpicke (2006a) noted that frequent self-testing during study may provide both direct and mediating benefits for long-term retention. A test may directly enhance retention through engaging those retrieval processes that will likely be important at the time of a later test (Roediger & Karpicke, 2006a). As an example of a mediating influence, students may use a test as feedback to ascertain what material requires further study and what material is already well known (Izawa, 1970, 1971; Thompson et al., 1978). The findings of the present study represent another example of a mediating influence of testing. Participants are better able to learn a new set of materials if previous information in the study sequence has been tested. We now relate these findings to the relevant literature and discuss their broader implications.

In each of our experiments, participants expected a final test. Accordingly, they were likely to hold words in mind (across lists) as they studied (Masson & McDaniel, 1981; Szpunar et al., 2007). Participants also knew that a test might follow the presentation of any list in the study sequence. Thus, although maintenance of all lists was important in the long term, subsequent lists in the study sequence had to be discriminated in anticipation of initial testing. Our results indicate that testing facilitated this discrimination process. Participants who had been tested while learning previous study lists (i.e., Lists 1–4) were better able to learn and retain a new study list (i.e., List 5), relative to participants who had not been previously tested. Specifically, tested participants were more adept at learning and discriminating the new study list from prior lists in the study sequence; they produced more correct responses and fewer prior-list intrusions.

According to the classic interference literature, the benefit of testing on list discrimination (in the context of a PI design) may be related to the enhanced retention of tested materials. Specifically, variables known to enhance long-term retention reduce the probability of PI (Melton, 1936; Postman, 1962; Underwood, 1949). For instance, Melton (1936) reported that overlearning a paired-associate word list (40 study–test trials) produced little interference on learning an ensuing list, relative to intermediate amounts of initial list learning. Underwood and Ekstrand (1967) demonstrated that spaced, relative to massed, practice of a paired-associate word list reduced the amount of interference on learning a subsequent list. The provision of testing during study may represent an extension of this general finding. Well-learned materials appear to be easily discriminated by participants as they acquire subsequent information.

Along the same lines, Postman and Keppel (1977) suggested that testing adds contextual elements (over and above those provided by a study episode; Anderson & Bower, 1972) that enhance the subsequent discriminability of recalled materials (cf. Chan & McDermott, 2007). That is, participants who engage in retrieval after study may have more (or more reliable) information upon which to base future discriminations (e.g., “I have recalled this word before”) than participants who must base future discriminations solely on study episodes (e.g., “Did I just learn this word, or did I see it in a previous list?”; cf. Jang & Huber, 2008).

We interpret our results in terms of the source-monitoring framework (Johnson, Hashtroudi, & Lindsay, 1993) and the reduction of cue overload (Watkins & Watkins, 1975). We propose that the benefit provided by testing after each list permits students to segregate lists from one another and to distinguish them, thus reducing cue overload during retrieval. That is, when students study and are tested on a fifth list when the first four have been studied but not tested, the functional search set is large relative to when testing of the four previous lists has occurred. We propose that when the four previous lists have been studied and tested, subjects find they can more easily circumscribe the relevant search set to the 18 recently presented words. However, when the first four lists have not been tested, subjects must try to recollect the most recent set of 18 words distinguished only by temporal cues. Thus, the functional search set may be larger (somewhere between 18 and 90 words) when the prior lists have not been tested, and, because of this, the probability of recall is reduced on the fifth list (due to cue overload) and the proportion of intrusion errors from prior lists is increased (due to a failure to delimit the set of items to be retrieved). Testing, then, serves as a powerful means of segregating the lists, permitting more efficient source monitoring and reduction of cue overload. This line of reasoning further provides a parsimonious interpretation of final recall performance. It has been well established that items that do not receive retrieval practice are less likely to be recalled at a later time (Karpicke & Roediger, 2008). Participants who had not been tested on Lists 1–4 were not able to initially recall many List 5 items. As a result, those items were also not recalled very well on the final test. Conversely, participants who had been tested on Lists 1–4 were able to initially recall a considerable number of List 5 items and subsequently incurred a greater benefit of retrieval for those items on the final test.

Some additional insight was gained in Experiment 3, which demonstrated that this benefit from reduced PI is specific to testing and not to the reexposure to material that accompanies a test. This result further emphasizes that a test likely provides unique contextual information upon which future discriminations can be based. An opportunity for repeated study clearly did not procure the same advantage. A test appears to produce fundamentally different consequences from additional study (Roediger & Karpicke, 2006b).

Finally, we address an apparent discrepancy between our findings and the classic interference literature in that reliable PI effects have been reported even after stimulus materials are repeatedly tested during study (Greenberg & Underwood, 1950; Postman & Keppel, 1977; Postman & Underwood, 1973; Underwood, 1957). One important distinction between our results and the classic interference literature is that previous PI studies rarely employed the use of a "no test" baseline condition (but see Tulving & Watkins, 1974). That is, although there likely exist many learning paradigms in which PI accumulates in the face of immediate testing, the present set of results suggests that the accumulation of PI would be greater in the absence of testing. Future work will need to evaluate this claim more carefully.

Our findings have important implications for study strategies. Specifically, in studying for a test, learners must invariably acquire a large set of materials. That is, one set of materials must be learned first, second, third, and so on. Here we show (using a laboratory analogue of this scenario) that learners may be vulnerable to buildup of PI. In fact, this is likely an experience that is common to anyone who has felt overloaded with information during the course of study. Our data directly address this experience. That is, although learners are likely to experience more difficulty in acquiring materials later in the study sequence, the negative influence of previous materials may be curbed through frequent recall of information. In terms of the classroom setting, frequent testing would benefit direct retention of the tested material and have two important indirect effects: PI in learning later material would be reduced and students would doubtless study more frequently.

References

Akers, H. R., & Lyons, D. J. (1979). A mediational explanation of the priority effect. *Journal of Verbal Learning and Verbal Behavior, 18,* 721–731.

Allen, G. A., & Arbak, C. J. (1976). The priority effect in the A-B, A-C paradigm and subjects' expectations. *Journal of Verbal Learning and Verbal Behavior, 15,* 381–385.

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review, 79,* 97–123.

Barnes, J. M., & Underwood, B. J. (1959). "Fate" of first-list associations in transfer theory. *Journal of Experimental Psychology, 58,* 97–105.

Chan, J. C. K., & McDermott, K. B. (2007). The testing effect in recognition memory: A dual process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 431–437.

Darley, C. F., & Murdock, B. B. (1971). Effects of prior free recall testing on final recall and recognition. *Journal of Experimental Psychology, 91,* 66–73.

Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior, 5,* 351–360.

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

Greenberg, R., & Underwood, B. J. (1950). Retention as a function of stage of practice. *Journal of Experimental Psychology, 40,* 452–457.

Indig, S. (2005, April 12). Cramming Canadians: Two-thirds of university students start studying for exams no more than a week in advance. Retrieved November 22, 2006, from http://www.kumon.com/pressroom/pressreleases/article-canada.asp?articlenum=17&language=Canada

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

Izawa, C. (1971). The test trial potentiating model. *Journal of Mathematical Psychology, 8,* 200–224.

Jang, Y., & Huber, D. E. (2008). Context retrieval and context change in free recall: Recalling from long-term memory drives list isolation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 112–127.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3–28.

Karpicke, J. D., & Roediger, H. L., III. (2008). The critical importance of retrieval for learning. *Science, 319,* 966–968.

Keppel, G. (1968). Retroactive and proactive inhibition. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory.* Englewood Cliffs, NJ: Prentice Hall.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Leeming, F. C. (2002). The exam-a-day procedure improves performance in psychology classes. *Teaching of Psychology, 29,* 210–212.

Malis, G. (1970). *The priority effect in the A-B, A-C paradigm.* Unpublished doctoral dissertation, University of Toronto.

Masson, M. E. J., & McDaniel, M. A. (1981). The role of organization processes in long-term retention. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 100–110.

Melton, A. W. (1936). Proactive inhibition in serial verbal learning as a function of the degree of learning of the first material. *Psychological Bulletin, 33,* 765.

Michael, J. (1991). A behavioral perspective on college teaching. *Behavior Analyst, 14,* 229–239.

Postman, L. (1962). Transfer of training as a function of experimental paradigm and degree of first-list learning. *Journal of Verbal Learning and Verbal Behavior, 1,* 109–118.

Postman, L., & Keppel, G. (1977). Conditions of cumulative proactive inhibition. *Journal of Experimental Psychology: General, 106,* 376–403.

Postman, L., & Underwood, B. J. (1973). Critical issues in interference theory. *Memory & Cognition, 1,* 19–40.

Robbins, D., & Irvin, J. R. (1976). The priority effect: Test effects on negative transfer and control lists. *Bulletin of the Psychonomic Society, 8,* 167–168.

Roediger, H. L., & 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., & Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17,* 249–255.

Sahakyan, L., Delaney, P. F., & Kelley, C. M. (2004). Self-evaluation as a moderating factor of strategy change in directed forgetting benefits. *Psychonomic Bulletin & Review, 11,* 131–136.

Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology, 30,* 641–656.

Szpunar, K. K., McDermott, K. B., & Roediger, H. L. (2007). Expectation of a final cumulative test enhances long-term retention. *Memory & Cognition, 35,* 1007–1013.

Thompson, C. P., Wenger, S. K., & Bartling, C. A. (1978). How recall facilitates subsequent recall: A reappraisal. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 210–221.

Tulving, E., & Psotka, J. (1971). Retroactive inhibition in free recall: Inaccessibility of information available in the memory store. *Journal of Experimental Psychology, 87,* 1–8.

Tulving, E., & Thornton, G. B. (1959). Interaction between proaction and retroaction in short-term retention. *Canadian Journal of Psychology, 13,* 255–265.

Tulving, E., & Watkins, M. J. (1974). On negative transfer: Effects of testing one list on the recall of another. *Journal of Verbal Learning and Verbal Behavior, 13,* 181–193.

Underwood, B. J. (1945). The effect of successive interpolations on retroactive and proactive inhibition. *Psychological Monographs, 59,* 1–33.

Underwood, B. J. (1949). Proactive inhibition as a function of time and degree of prior learning. *Journal of Experimental Psychology, 39,* 24–34.

Underwood, B. J. (1957). Interference and forgetting. *Psychological Review, 64,* 49–60.

Underwood, B. J., & Ekstrand, B. R. (1967). Studies of distributed practice: XXIV. Differentiation and proactive inhibition. *Journal of Experimental Psychology, 74,* 574–580.

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language, 50,* 289–335.

Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory, 104,* 442–452.

Wickens, D. D. (1970). Encoding categories of words: An empirical approach to meaning. *Psychological Review, 77,* 1–15.

Wickens, D. D. (1972). Characteristics of word encoding. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory.* New York: Wiley.

Zaromb, F. M., Howard, M. W., Dolan, E. D., Sirotin, Y. B., Tully, M., Wingfield, A., et al. (2006). Temporal associations and prior-list intrusions in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 792–804.

Received February 23, 2008
Revision received May 29, 2008
Accepted May 29, 2008 ■