Memory & Cognition
2007, 35 (5), 1007-1013

Expectation of a final cumulative test enhances long-term retention

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

In the present study, subjects studied lists of words across four experimental conditions: whether (or not) subjects received initial testing for these lists and whether (or not) they were made aware of an upcoming final free-recall test. Initial testing enhanced final-test performance; however, subjects benefited more from initial testing when they also knew they would need to remember the information for a later test. The data suggest that holding an expectation of the final test encouraged the continued processing of study materials following an initial test, affecting the accessibility of these materials at the time of final recall. The results clearly illustrate how an expectation of a cumulative test might influence long-term retention, which may have important implications for educational practice.

About a century ago (see, e.g., Abbott, 1909; Gates, 1917), researchers began to consider that learning could take place when information was retrieved from memory (see also Lachman & Laughery, 1968; Spitzer, 1939; Tulving, 1967). Since then, they have accumulated considerable evidence showing that intermediate testing between study and final recall enhances retention (see, e.g., Allen, Mahler, & Estes, 1969; Darley & Murdock, 1971; Hogan & Kintsch, 1971; Masson & McDaniel, 1981; McDaniel & Masson, 1985; for a review, see Roediger & Karpicke, 2006). In most such studies, the final test is presented incidentally, to ensure that subjects do not continue processing study materials after the initial test. This design feature ensures that researchers arrive at accurate estimates of the influence of intermediate tests on final recall. However, might this experimental parameter work against the subject? As discussed below, the present study was designed to examine whether the long-term retention of previously tested information is influenced by the expectation of a final test.

We develop here a description of how the expectation of a final test might influence long-term retention. We refer to this as the *expectation hypothesis*.[1] After taking an initial test, subjects expecting a final test know that the information initially tested will be relevant at a later time and that they should attempt to continue processing that information. In contrast, subjects not expecting a final test should have no reason to engage in further, additional processing; taking the initial test should act as a signal that continued processing of the material is not required. To the extent that subjects expecting a final test are able to make effective use of additional processing of previously tested materials, the benefits of initial testing (Glover, 1989; Spitzer, 1939) should be enhanced by the expectation of future testing.

The notion that an expectation of a final test may influence long-term retention has been largely ignored. Shimizu (1996), however, did address the issue while evaluating the effectiveness of various rehearsal strategies on final-test performance. In his study, subjects learned five lists of unrelated words. Studying each list was followed by an immediate recall test. One variable of interest was whether subjects expected a final free-recall test. The other was whether subjects were constrained to use a specific rehearsal strategy or were allowed to use a strategy of their own choosing. Of interest for the present study, Shimizu found no evidence of a final-test expectancy effect when strategy choice was restricted. An expectancy effect did appear when subjects were free to rehearse as they pleased, but this outcome appears to have been an artifact of initial test performance. That is, the group expecting the final test outperformed the unsuspecting group in *both* initial and final recall (see Shimizu, 1996, Figure 4); there was no difference in the amount of information retained from initial to final testing between the two groups of subjects. Thus, although the author claimed to show an expectancy effect resulting from subjects' intentional shifts in strategies, the data do not clearly bear this out.

This finding is at odds with the expectation hypothesis we propose, which would predict superior long-term retention on the part of subjects expecting a final test. Closer consideration of the specific conditions of Shimizu's (1996) experiment is instructive. The materials used were five lists of 10 unrelated words. Although subjects expecting to take a final test knew that it would be beneficial to keep previously tested information in mind, they may have been unable to do so. Since none of the words within any list were related, subjects may have had difficulty relating previously tested items to words in subsequent lists. Simply being aware of the final test may have done little if subjects were unable to take advantage of the opportunity to further process previously tested materials (see also

K. K. Szpunar, karl.szpunar@wustl.edu

Runquist, 1983). The present experiment was designed to examine this issue. We were interested in both the benefit of initial testing and whether this benefit might be affected by the expectation of a final test. We varied whether subjects received intermittent testing on a series of five separate word lists, culminating in a final free-recall test. We also varied whether subjects were made aware of the final cumulative test. Most importantly, study materials were designed to allow for posttest processing. That is, words were semantically related *both* within and across lists.

An additional goal of the present experiment was to test our hypothesis that subjects treat a test as a cue to forget (i.e., to cease rehearsal) when a final test is not expected (cf. MacLeod, Dodd, Sheard, Wilson, & Bibi, 2003). To that end, a separate group of subjects was both unaware that a final test would occur and was instructed to immediately forget everything they had learned after having taken each initial test.

METHOD

Subjects

One hundred sixty Washington University undergraduates participated in the experiment as a course requirement. The subjects were tested in small groups that were randomly assigned to one of four experimental conditions, resulting in 40 subjects per condition.

Materials

We constructed five 18-word lists. 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 as follows: building parts, earth formations, animals, fruits, human body parts, and weather phenomena. In constructing the lists, the first five exemplars from each category were omitted. This was done to reduce the likelihood of subjects' correctly guessing words, since these first exemplars are most likely to be highly associated with their respective categories. The next 15 exemplars (6–20) in each category were then divided into five groups of three words each. These triads were then assembled into five separate word lists. Experimental materials were presented using E-Prime software (Psychology Software Tools).

Design

Each of four groups of subjects studied a series of five separate word lists. List and word order were randomized across subjects. The two manipulations of interest were whether subjects received initial testing for each of the five word lists and whether they were made aware of an upcoming final free-recall test.

Procedure

All subjects were initially informed that the experiment was designed to test their memory and mathematical ability. They were also told that five separate word lists would be presented to them visually, in the center of a computer monitor. Each word in the succession was presented for 2 sec (500-msec interstimulus interval, 2,500-msec stimulus onset asynchrony), and subjects were instructed to pay close attention to each word.

Subjects who were initially tested and aware of the final test—referred to henceforth as the *aware* group—were told that they would complete a set of two tasks following the presentation of each list. First, they were to solve math problems for 1 min, in order to minimize primary memory effects in free recall (Glanzer & Cunitz, 1966). Then they were given 1 min to recall, in any order, as many words as they could from the list they had studied just prior to having solved the math problems. All responses—math problems and list recall—were typed using a computer keyboard and were clearly visible on the computer screen. Prior to beginning the experiment, these subjects were made aware of a final 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, subjects 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 subjects were to use all 8 min efficiently in their attempt to recall study materials. They were also instructed to draw a line every minute under the last word they had recalled, which would permit us to construct cumulative recall curves.

Subjects in the *untested control* group were also made aware of the final test. They were not, however, initially tested on each of the five separate word lists. Instead, after studying each list and completing the numerical exercise that followed, the subjects simply completed a new set of mathematical problems. Adding the second numerical exercise resulted in equal time between presentations of word lists across all four experimental conditions. The remainder of the procedure was left unchanged.

Subjects who were initially tested and unaware of the final test—the *unaware* group—received instructions similar to those given to the aware group and completed similar tasks, with one exception: At the outset of the experiment, these subjects were not alerted to the fact that there would be a final test. That is, they completed each of the five separate study–test segments as mentioned before; they then received the surprise final test after the 30-min retention interval.

Finally, subjects who were initially tested, unaware of the final test, and explicitly cued to forget each list—the *unaware–cue* group—completed the experimental tasks in exactly the same fashion as did subjects in the unaware group, but with one additional set of instructions: This group was instructed to consciously forget the list they had just studied and recalled. For each of the five lists, the study–test–forget cycle was completed before subjects moved on to the next list. These subjects were told that actively ignoring past information would help them maintain a high level of performance across all five of the word lists.

All subjects completed the experiment in approximately 1 h. They were then informally interviewed about the experimental session. Specifically, we wanted to know whether subjects had noticed the relations of words across lists and, more important, what their reaction was to having noticed these relations. After this brief interview, we thanked all subjects for their participation and they were fully debriefed.

RESULTS AND DISCUSSION

Initial Tests

For each subject, we calculated a composite initial-test score based on the combined performance across all five initial tests. Means for the three condition groups receiving initial testing—aware, unaware, unaware–cue—are presented in the left panel of Figure 1. Because these means are based on performance across all five initial tests, a 3 (condition) $\times$ 5 (test) repeated measures ANOVA was applied to the initial-test data. This analysis revealed no significant main effects or interaction, indicating that all groups had performed comparably across all five initial tests; in fact, all groups initially recalled approximately 50% of each study list.

We hypothesized that subjects in the two unaware groups should not have attempted to retain previously tested materials as they proceeded through the study; it is, therefore, surprising that these subjects did not exhibit higher overall initial-test performance than subjects in the aware group. That is, one might have expected subjects in the aware group to experience greater proactive interference, since they were attempting to keep previously tested information in mind for the final test. However, the data do not show any evidence of this proactive interference. The design of the present study provides some insight into the nature of this outcome. Specifically, the related nature of the words—both within and across lists—may have boosted performance for all groups, obviating any effect of whether information was being held in mind (see, e.g., Horton & Petruk, 1980; Woodward & Bjork, 1971). This null outcome on the initial tests still does not preclude finding interesting differences among these groups in final-recall performance.

Final Tests

Mean final-recall performance is presented in the right panel of Figure 1. A one-way ANOVA revealed a significant main effect of condition [$F(3,156) = 29.24$, $MS_e = .014$, $p < .001$]. Upon further investigation, independent-samples *t* tests revealed that all groups receiving initial testing—aware, unaware, and unaware–cue—recalled significantly more words on the final test than did the untested group [$ts(78) = 9.53$, 6.98, and 5.91, respectively, $ps < .001$]. In addition, the aware group recalled significantly more words on the final test than did the unaware group [$t(78) = 2.50$, $p = .015$] and the unaware–cue group [$t(78) = 2.26$, $p = .026$]. The two unaware groups did not differ reliably from one another. Two interesting patterns of results emerged from our data—one related to the influence of testing and the other related to the influence of expecting a final test. The remainder of our analyses focus on these two patterns of results and aim to elucidate the nature of their origins.

Testing

All three groups receiving initial testing recalled significantly more words on the final test than did the untested group. This outcome was expected and reinforces the well-documented finding that testing has positive benefits for long-term retention (for a review, see Roediger & Karpicke, 2006). To gain further insight into the nature of this outcome, we partitioned words recalled on the final test into their origins of study. That is, we were interested in examining which of the five original study lists contributed to words recalled on the final test and whether there were any differences across conditions. This analysis was motivated by a previous study (Tulving & Thornton, 1959).

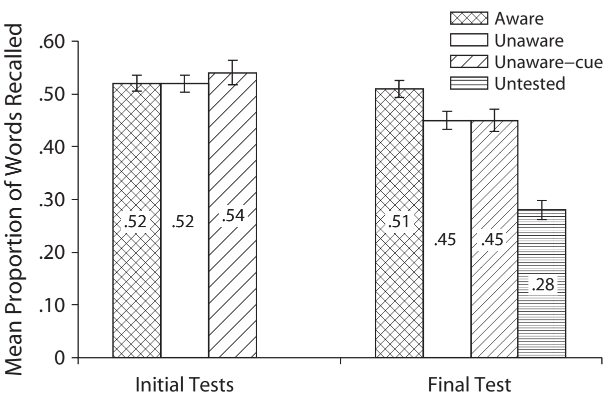

Figure 1. Mean proportions of study words recalled during initial testing and the final test as a function of condition. Error bars display standard errors of the mean.

Figure 2 illustrates the partitioning of the final free-recall test data according to list of study. It is apparent that—along with demonstrating better overall performance—tested groups (relative to the untested group) demonstrated a marked effect of recency in recall of originally studied lists. Recall was better for Lists 4 and 5 ($M = .51$) than for Lists 1 and 2 ($M = .42$). In contrast, the untested group exhibited an effect of primacy, with recall better for the first list (.37) relative to later lists (.25). Confirming this apparent relationship between testing and resulting pattern of recall, a 2 (group) $\times$ 5 (list) repeated measures ANOVA revealed a significant group $\times$ list interaction [$F(4,632) = 14.43$, $MS_e = .019$, $p < .001$]. To verify that the observed effects of recency and primacy were indeed real, we compared mean List 1 recall from the final test to mean List 5 recall from the final test for each group. The tested group, as a whole, showed significantly better recall of the more recently learned list words [$t(119) = -5.15$, $p < .001$]. The untested group demonstrated a recall advantage for list words presented in the earliest stages of study [$t(39) = 3.79$, $p < .001$]. Further inspection of Figure 2 suggests that the testing effect in final recall was apparent only in Lists 2–5. Statistical tests confirmed this observation, showing that the initially tested subjects recalled reliably more words than the initially untested subjects for study Lists 2, 3, 4, and 5 [$ts(158) = 5.93$, 7.34, 8.52, and 7.93, respectively, $ps < .001$], but not for List 1 ($p > .20$).

Expectancy of a Final Test

Comparisons of mean final-recall performance also revealed significant differences among groups receiving initial testing. Specifically, the aware group recalled more words on the final test than did either of the unaware groups. To gain further insight into this obtained effect of expectancy, a variant of trial-to-trial recall analysis (Tulving, 1964; see also Estes, 1960) was conducted for all three groups that had received both initial and final testing. Such an analysis is typically concerned with conditionalizing recall performance in one trial (recalled vs. not) based on recall performance from a previous trial (recalled vs. not). This allows one to partition performance on a given test beyond the simple dichotomy of *recalled versus not recalled* into a set of four conditional probabilities. These include the complementary probabilities of recalling (CC) or not recalling (CN) a word on a subsequent test, given that it had previously been recalled, along with the complementary probabilities of recalling (NC) or not recalling (NN) a word on a subsequent test, given that it had not been previously recalled. In the present study, this analysis was accomplished by treating all five initial tests as test *n* and the final test as test $n+1$. This analysis provides a more fine-grained understanding of

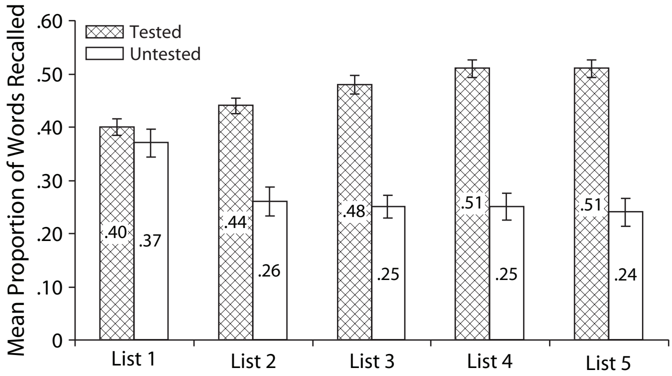

Figure 2. Final recall performance as a function of original study list and condition. Error bars display standard errors of the mean.

the obtained difference between the aware group and the two unaware groups.

Figure 3 compares intertest forgetting (CN) and reminiscence (NC) across the three groups of interest, given that the two remaining conditional probabilities were complementary. The aware group was less likely to forget words that had initially been recalled (CN) than were the unaware group [$t(78) = 3.26, p = .002$] and the unaware–cue group [$t(78) = 3.79, p < .001$]. The aware group was also more likely to recover initially missed words (NC) than were the unaware group [$t(78) = 2.11, p = .038$] and the unaware–cue group [$t(78) = 2.42, p = .018$]. The two unaware groups did not differ from each other in either respect. Considering the comparable performance of the three groups in initial testing, these novel results support our hypothesis that the expectation of a final test may enhance long-term retention.

Cumulative Recall Curves

We present here a set of analyses that highlight the unique influences of testing and expectancy. At first glance, prior testing and the expectation of a final test appear to affect final-recall performance to quantitatively different degrees, with testing having a greater impact on long-term retention than expectancy (see Figure 1). However, closer examination of the final free-recall data reveals a more interesting story.

Subjects were allotted 8 min for the final free-recall test. They were instructed to draw a line every minute under the last word they had recalled, allowing us to construct cumulative recall curves, as shown in Figure 4. It is apparent that—along with having a greater impact on final-recall performance—testing imparted its influence sooner in final recall than did the expectation of a final test. Confirming this observation, statistical analyses revealed that the effect of testing first became apparent after only 1 min of recall, the point at which the number of words recalled by the aware, unaware, and unaware–cue groups had already reliably exceeded that recalled by the untested group [ts$(78) = 6.25, 4.99$, and 5.13, respectively, ps $< .001$]. The effect of expectancy did not manifest itself until the second minute of recall, when the number of words recalled by the aware group reliably exceeded that of the unaware group [$t(78) = 2.64, p = .010$] and the unaware–cue group [$t(78) = 2.24, p = .028$]. These results emphasize the powerful impact of testing on long-term retention and show that the effect of testing may be enhanced by the expectation of a future test—at least in due time.

GENERAL DISCUSSION

We examined the benefits of prior testing on long-term retention and whether these benefits may be affected by an expectation of a final test. Our findings are notable in at least four respects: (1) Taking an initial test enhanced performance on the final test, relative to not having taken an initial test. (2) Having an expectation of a final test enhanced final-test performance, relative to not having that expectation. (3) An explicit cue to forget was shown to be unnecessary in producing the observed difference between subjects expecting a final test and those not expecting that test. (4) Conditional analyses revealed that the expectation of a final test kept studied words in a more accessible state, reducing forgetting between initial and final testing and aiding reminiscence. We now relate these findings to the relevant literature and discuss their broader implications.

The finding that taking an initial test enhanced final-test performance was expected and serves simply to replicate the long-standing knowledge that testing enhances long-term retention. Of interest, however, was the breakdown

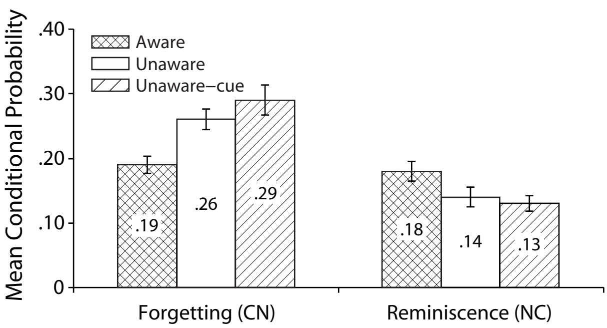

Figure 3. Mean conditional probabilities of forgetting and reminiscing for groups receiving both initial and final testing. Error bars display standard errors of the mean.

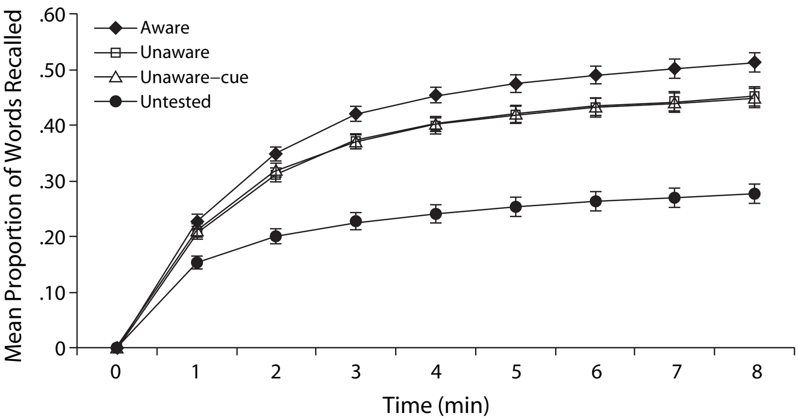

Figure 4. Cumulative recall curves as a function of condition. Error bars display standard errors of the mean.

of words recalled in the final test in terms of their origin in the original study lists (see Figure 2). This analysis—comparing all initially tested groups with an initially untested group—revealed that initial testing was associated with increasingly better recall throughout the list sequence (a list recency effect). In contrast, the absence of initial testing was associated with a recall advantage for list words presented in the earliest stages of study (a list primacy effect).

The recency advantage in recall of lists for which initial testing had been administered (cf. Zimmerman & Underwood, 1968, Figure 1) has previously been attributed to the influence of retroactive interference—the learning of later lists interfering with the retention of earlier lists—at the time of final recall (Tulving & Thornton, 1959). Consonant with the present study, Tulving and Thornton observed no difference in the recall of study lists during initial testing and a marked effect of recency emerging in final free recall. In a thorough investigation of the phenomenon—in which they manipulated the degree to which list study was subjected to proactive and/or retroactive interference—Tulving and Thornton found no evidence of proactive interference affecting final recall performance. They did, however, find a significant influence of retroactive interference (see also Tulving & Psotka, 1971; for an alternative explanation, see Sahakyan, Delaney, & Kelley, 2004).

As for the untested group—a manipulation not examined by Tulving and Thornton (1959)—the present study provides the first evidence of primacy in recall in the context of a multilist learning experiment (but see Tulving & Watkins, 1974, for a similar outcome in paired-associate learning). Related research points to the possibility that this observed result may be due to the influence of proactive interference. For instance, Darley and Murdock (1971) showed that words from previously untested lists were more likely than words from lists that had been tested earlier to intrude upon the recall protocols of subsequently tested lists. This observation—along with the finding that groups that had been initially tested showed comparable performance for all lists (in initial testing)—suggests that the act of testing may serve a protective function against the influence of proactive interference. This may help explain why the untested group displayed such poor retention of materials following the first list, whereas the tested groups showed initially proportionate learning of all lists.

Of particular interest, we found that, given initial testing, holding an expectation of a final test led to better performance on that test, as opposed to not holding that expectation. A previous study (Shimizu, 1996; see also Runquist, 1983) using unrelated word lists also hinted at this effect. Our use of words that were related both within and across lists facilitated additional processing of previously tested materials; this feature of our design explains why our pattern was more pronounced. Also, the performance patterns of the two groups not expecting the final test were shown not to differ, suggesting that an explicit instruction to forget was not necessary in order to produce the effect (MacLeod et al., 2003; see also Epstein, 1969a, 1969b, 1970). That is, simply removing the expectation of a cumulative test led subjects to cease further processing of previously tested materials, leading to the reduced accessibility of those materials at the time of final recall. In effect, not having an expectation of a final test may operate like an instruction to forget.

Holding an expectation of future testing led to less forgetting and more item recovery between the initial and final test, suggesting that holding the expectation of testing influenced the accessibility (Tulving & Pearlstone, 1966) of the study set as a whole at the time of final recall. Why should this be so? As outlined by the expectation hypothesis, we propose that the extent to which subjects engaged in continued processing of study materials after an initial test differed between subjects expecting a final test and those not expecting it. Specifically—given that the final test was cumulative—subjects expecting it should have been more inclined to keep previously learned materials in mind as they proceeded through the study. As a result, they may have been more likely to notice the relations of words across lists, in effect setting up strong associations across all study materials (see Masson & McDaniel, 1981). Interrelating list items could easily have occurred due to the overlap of categories across lists (Rundus, 1971, Experiment 4). In contrast, a subject not expecting a final test would have no incentive to pay particular attention to such relations across lists and should, therefore, benefit less from the associations. Subject reports across all conditions suggested that previous words were, in fact, coming to mind during the study of subsequent lists. In line with our interpretation, subjects expecting the final test reported consciously taking notice of these relations across lists, whereas subjects not expecting the final test reported exerting effort in concentrating their attention on the list they were presently learning.

If our interpretations of the data are correct, one should expect these proposed differences in the processing of study words after an initial test to manifest themselves at the time of final recall. Indeed, all subjects were found to cluster their recall by category in the final test (see, e.g., Bousfield & Bousfield, 1966; for a review, see Schuell, 1969), indicating that awareness of the relations of words across lists influenced later recall. Given that subjects in the aware group had reported paying particular attention to these relations—and that subjects in the two unaware groups had no reason to do so—we suspected that subjects expecting the final test might have clustered their output to a greater degree than subjects not expecting the test. An examination of the final-recall protocols (see Table 1)—using the adjusted-ratio-of-clustering method, which controls for differences in level of recall (Roenker, Thompson, & Brown, 1971) revealed that subjects in the aware group were, in fact, substantially more likely to cluster their output than subjects in the unaware group [$t(78) = 4.12, p < .001$] and in the unaware–cue group [$t(78) = 3.70, p < .001$]. Again, data from the two unaware groups did not differ from one condition to the other [$t(78) < 1$]. These findings nicely complement subjective reports. Anticipation of the final test led to greater benefits procured from noticing the relations of words across lists, perhaps because holding an expectation of the final test encouraged seeking some relations among study materials throughout the course of the experiment.

The idea that a subject's anticipation of experimental parameters should influence long-term retention is not new (Müller, 1911). In fact, our interpretation of the present data bears a strong resemblance to that of prior studies. For example, Jacoby and Bartz (1972) examined the role of rehearsal in transfer to long-term memory and suggested that the manner in which list words were processed—particularly those in terminal positions—depended on the anticipation of the type of interval between study and test. Terminal list words were better recalled during initial testing when study was followed by either an unfilled interval or immediate testing than after a filled delay. However, the opposite was shown to be true at the time of a final free-recall test. The authors (see also Götz & Jacoby, 1974; Jacoby, 1973; Meunier, Ritz, & Meunier, 1972; Roenker, 1974) argued that the expectation of a filled interval led to more elaborate processing (see, e.g., Craik & Lockhart, 1972; Woodward, Bjork, & Jongeward, 1973) of those words, in comparison with an unfilled or no interval, which required subjects only to maintain those words until recall (e.g., Craik, 1970; Craik, Gardiner, & Watkins, 1970; Watkins & Watkins, 1974). As a result, terminal list words following a filled interval were better retained in the long run because of the more elaborate processing during study.

Table 1
Mean Adjusted-Ratio-of-Clustering (ARC) Scores and Standard Deviations for Groups Receiving Both Initial and Final Testing

Condition	ARC *M*	ARC *SD*
Aware	.66	.16
Unaware	.51	.17
Unaware–cue	.52	.23

Note—Range for ARC = 0–1, where 0 represents *no clustering* and 1 represents *perfect clustering*.

Likewise, subjects expecting the final test in the present study appeared to use their knowledge of an upcoming test to guide the manner in which they processed study words across all lists, in effect setting up relations that would enhance final recall performance. Conversely, subjects not expecting the final test had no reason to pay attention to the relations of words across lists and, therefore, focused their attention on learning each list individually, moving on to each successive list under the impression that anything they had studied and recalled before would no longer be relevant.

In the typical classroom setting, material is taught in the form of units, each of which is usually followed by a test—a situation similar to the one constructed in the present experiment. Likewise, some courses require a cumulative final examination, and others do not. Students expecting a cumulative final exam may attempt to relate information across units as they learn. On the other hand, students not expecting a cumulative exam might feel inclined to disregard any previously tested information and focus their attention only on new information. Although the parameters of the present study—five lists of 18 words, a 1-min delay between study and initial test, and a 30-min delay between initial and final testing—are certainly not representative of those of the classroom, it seems plausible that the present results may indicate that students study more effectively when they expect a final examination. Of course, the effect in actual courses may be even greater than the one estimated here, because, in courses, students may use additional study of earlier presented material in preparing for the final exam. Our study shows that—even without extra overt study—expectation of a final exam enhances performance.

AUTHOR NOTE

This research was supported by a grant from the James S. McDonnell Foundation (220020041) to the second and third authors. The research was conducted as part of the first author's master's thesis and was presented as a poster at the 17th Annual Convention of the American Psychological Society in Los Angeles, CA. The manuscript benefited greatly from comments provided by Jason Chan, Sean Kang, Jessica Logan, and Mark McDaniel. Correspondence concerning this article may be sent to K. K. Szpunar, Department of Psychology, Campus Box 1125, Washington University, One Brookings Drive, St. Louis, MO 63130 (e-mail: karl.szpunar@wustl.edu).

REFERENCES

Abbott, E. E. (1909). On the analysis of the factors of recall in the learning process. *Psychological Monographs*, **11**, 159-177.

Allen, G. A., Mahler, W. A., & Estes, W. K. (1969). Effects of recall tests on long-term retention of paired associates. *Journal of Verbal Learning & Verbal Behavior*, **8**, 463-470.

Bousfield, A. K., & Bousfield, W. A. (1966). Measurement of clustering and of sequential constancies in repeated free recall. *Psychological Reports*, **19**, 935-942.

Craik, F. I. M. (1970). Fate of primary memory items in free recall. *Journal of Verbal Learning & Verbal Behavior*, **9**, 143-148.

Craik, F. I. M., Gardiner, J. M., & Watkins, M. J. (1970). Further evidence for a negative recency effect in free recall. *Journal of Verbal Learning & Verbal Behavior*, **9**, 554-560.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior*, **11**, 671-684.

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.

Epstein, W. (1969a). Poststimulus output specification and differential retrieval from short-term memory. *Journal of Experimental Psychology*, **82**, 168-174.

Epstein, W. (1969b). Recall of word lists following learning of sentences and of anomalous and random strings. *Journal of Verbal Learning & Verbal Behavior*, **8**, 20-25.

Epstein, W. (1970). Facilitation of retrieval resulting from postinput exclusion of part of the input. *Journal of Experimental Psychology*, **86**, 190-195.

Estes, W. K. (1960). Learning theory and the new "mental chemistry." *Psychological Review*, **67**, 207-223.

Gates, A. I. (1917). Recitation as a factor in memorizing. *Archives of Psychology*, **6**, No. 40.

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

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

Götz, A., & Jacoby, L. L. (1974). Encoding and retrieval processes in long-term retention. *Journal of Experimental Psychology*, **102**, 291-297.

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

Horton, K. D., & Petruk, R. (1980). Set differentiation and depth of processing in the directed forgetting paradigm. *Journal of Experimental Psychology: Human Learning & Memory*, **6**, 599-610.

Jacoby, L. L. (1973). Encoding processes, rehearsal, and recall requirements. *Journal of Verbal Learning & Verbal Behavior*, **12**, 302-310.

Jacoby, L. L., & Bartz, W. H. (1972). Rehearsal and transfer to LTM. *Journal of Verbal Learning & Verbal Behavior*, **11**, 561-565.

Lachman, R., & Laughery, K. R. (1968). Is a test trial a training trial in free recall learning? *Journal of Experimental Psychology*, **76**, 40-50.

Lundeberg, M. A., & Fox, P. W. (1991). Do laboratory findings on test expectancy generalize to classroom outcomes? *Review of Educational Research*, **61**, 94-106.

MacLeod, C. M., Dodd, M. D., Sheard, E. D., Wilson, D. E., & Bibi, U. (2003). In opposition to inhibition. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 43, pp. 163-214). San Diego: Academic Press.

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

McDaniel, M. A., & Masson, M. E. J. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 371-385.

Meunier, G. F., Ritz, D., & Meunier, J. A. (1972). Rehearsal of individual items in short-term memory. *Journal of Experimental Psychology*, **95**, 465-467.

Meyer, G. (1934). An experimental study of the old and new types of examination: I. The effect of the examination set on memory. *Journal of Educational Psychology*, **25**, 641-661.

Meyer, G. (1936). The effect of recall and recognition on the examination set in classroom situations. *Journal of Educational Psychology*, **27**, 81-99.

Müller, G. E. (1911). Zur Analyse der Gedachtnistatigkeit und des Vorstellungsverlaufes: Teil I. [On the analysis of memory activity and the conception process: Part I]. *Zeitschrift für Psychologie*, Suppl. 5.

Roediger, H. L., III, & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, **1**, 181-210.

Roenker, D. L. (1974). Role of rehearsal in long-term retention. *Journal of Experimental Psychology*, **103**, 368-371.

Roenker, D. L., Thompson, C. P., & Brown, S. C. (1971). Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin*, **76**, 45-48.

Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, **89**, 63-77.

Runquist, W. N. (1983). Some effects of remembering on forgetting. *Memory & Cognition*, **11**, 641-650.

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.

Schuell, T. J. (1969). Clustering and organization in free recall. *Psychological Bulletin*, **72**, 353-374.

Shimizu, H. (1996). Rehearsal strategies, test expectancy, and memory monitoring in free recall. *Memory*, **4**, 265-287.

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

Tulving, E. (1964). Intratrial and intertrial retention: Notes towards a theory of free recall verbal learning. *Psychological Review*, **71**, 219-237.

Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning & Verbal Behavior*, **6**, 175-184.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning & Verbal Behavior*, **5**, 381-391.

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 & Verbal Behavior*, **13**, 181-193.

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 & Language*, **50**, 289-335.

Watkins, M. J., & Watkins, O. C. (1974). Processing of recency items for free recall. *Journal of Experimental Psychology*, **102**, 488-493.

Woodward, A. E., & Bjork, R. A. (1971). Forgetting and remembering in free recall: Intentional and unintentional. *Journal of Experimental Psychology*, **89**, 109-116.

Woodward, A. E., Bjork, R. A., & Jongeward, R. H. (1973). Recall and recognition as a function of primary rehearsal. *Journal of Verbal Learning & Verbal Behavior*, **12**, 608-617.

Zimmerman, J., & Underwood, B. J. (1968). Ordinal position knowledge within and across lists as a function of instructions in free-recall learning. *Journal of General Psychology*, **79**, 301-307.

NOTE

1. In order to avoid any confusion, it should be noted that there already exists a separate literature on test expectancy related to the *type* of test a subject expects to receive (e.g., Meyer, 1934, 1936; for a review, see Lundeberg & Fox, 1991). This phenomenon is different from the one of interest in the present report.

(Manuscript received March 9, 2006;
revision accepted for publication June 2, 2006.)