

Running Head: EXPECTATION OF LONG-TERM RETENTION

Expectation of a Final Cumulative Test Enhances Long-term Retention

Karl K. Szpunar, Kathleen B. McDermott, and Henry L. Roediger III

Washington University in St. Louis

Address Correspondence to:

Karl K. Szpunar

Department of Psychology

Washington University

St. Louis, MO 63130

Phone: (314) 935-8892

Email: karl.szpunar@wustl.edu

Abstract

In the present study subjects studied lists of words, and we varied (across four experimental conditions) whether or not they 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 (relative to no initial tests); however, subjects benefited more from initial testing if they 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 the nature by which an expectation of a cumulative test might influence long-term retention, and may have important implications for educational practice.

Expectation of a Final Cumulative Test Enhances Long-term Retention

One hundred years ago (e.g., Abbott, 1909; Gates, 1917), researchers began to consider that learning can take place when information is retrieved from memory (see also Lachman & Laughery, 1968; Spitzer, 1939; Tulving, 1967). Since then, considerable evidence has accumulated showing that an intermediate test between study and final recall enhances retention, as compared with a similar scenario in which an intermediate test is not administered (e.g., Allen, Mahler, & Estes, 1969; Darley & Murdock, 1971; Hogan & Kintsch, 1971; Masson & McDaniel, 1981; McDaniel & Masson, 1985; for a review see Roediger & Karpicke, in press). In most such studies, the final test is presented incidentally, so as to ensure that subjects do not continue to process study materials after the initial test. This design feature assures researchers a precise estimate of the influence of the intermediate test on final recall. However, might this experimental parameter work against the subject? As discussed below, the present study set out 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, which we refer to 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 should in some way attempt to continue processing that information. In contrast, subjects not expecting a final test should have no reason to engage in similar processing; taking the initial test should act as a signal that any further processing of the material is no longer required. To the extent that subjects expecting a final test are able to make effective use of any additional processing of

previously-tested materials, the benefits of testing (Glover, 1989; Spitzer, 1939) should be enhanced by the expectation of future testing.

While the notion that an expectation of a final test may influence long-term retention has been largely ignored, Shimizu (1996) did address the issue in the context of evaluating the effectiveness of various rehearsal strategies on final test performance. In his study, subjects learned five lists of unrelated words, each of which was followed by an immediate recall test. The two variables of interest were whether or not subjects expected a final free recall test and whether they were constrained to using a specific rehearsal strategy or allowed to use a strategy of their own choosing. Of interest to the present study, Shimizu found no evidence of a final test expectancy effect when strategy choice was restricted. Such an effect did appear when subjects were free to rehearse as they pleased, but this outcome appears to be 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 proposed by the present authors, which would predict superior long-term retention on the part of subjects expecting the final test. Closer consideration of the particular conditions of Shimizu's (1996) experiment is instructive. The materials used were five short lists of ten unrelated words. Although subjects expecting the final test knew that keeping previously-tested information in mind would be beneficial, they might have been unable to do so; because

all list words were unrelated, subjects may have had difficulty relating previously-tested items to later list words, while at the same time trying to learn the new unrelated words for the upcoming test. Hence, simply having kept the final test in mind may have done little if subjects were unable to take advantage of the opportunity to further process previously-tested materials (see also Runquist, 1983).

The present experiment was designed to examine this issue. Specifically, 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 for a series of five separate word lists leading to a final free recall test and whether or not subjects were made aware of the final cumulative test. Importantly, study materials were designed to afford post-test 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., cease rehearsal) when a final test is not expected (cf. MacLeod et al., 2003). To that end, a separate group of subjects was both unaware that a final test would occur and instructed to forget everything they had learned earlier, just after taking each initial test.

Method

Subjects. One hundred and sixty Washington University undergraduates participated in the experiment for partial fulfillment of a course requirement. The subjects were tested in small groups that were assigned at random 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 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. 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 during testing, as these high associates are most likely to be produced from their respective categories. The next 15 exemplars (6-20) in each category were then divided into five groups of three words each. These category triads were then assembled to form the five separate word lists. Experimental materials were presented using E-Prime software.

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 or not subjects received initial testing for each of the five word lists and whether or not 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 further told that five separate word lists would be visually presented to them in the center of a computer monitor. The study words were presented at two seconds per word (500ms inter-stimulus interval, 2500ms SOA), and subjects were instructed to pay close attention to each word.

Subjects who were initially tested and aware of the final test, referred to from henceforth as the *aware* group, were instructed that they would complete a set of two tasks following the presentation of each study list. First, they solved math problems for one minute to minimize primary memory effects in free recall (Glanzer & Cunitz, 1966).

Then they were given one minute to recall, in any order, as many words as possible from the list they had just studied. 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 fully aware of a final free recall test that would be administered approximately 30 minutes after completion of the initial five study-test segments. The retention interval involved participation in an unrelated verbal exercise. During the final test eight minutes were allotted for subjects to recall as many words as they could from all five lists of words they had studied. Responses were written with paper and pencil. It was stressed that they were to use all eight minutes 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 (in order to permit cumulative recall curves to be constructed).

Subjects in the *untested control* group were also made aware of the final test. However, they were not initially tested on each of the five separate word lists. Instead, after studying each list and completing the numerical exercise that followed, the exercise was simply repeated with a new set of mathematical problems. Adding the second numerical exercise equated time between presentations of word lists across 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 similar instructions to those of the aware group and completed similar tasks, with one exception. At the outset of the experiment they 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 and then received the surprise final test after the 30-minute 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 experimental tasks in exactly the same fashion as subjects in the unaware group with one additional set of instructions. This group was instructed to actively forget each list that they had just studied and recalled before they moved onto the next study-test segment. These subjects were told that actively ignoring past information would help maintain a high level of performance across all subsequent word lists.

The experiment was completed in approximately one hour by all subjects, at which point they were informally interviewed about the experimental session. Specifically, we were interested in knowing whether subjects had noticed the relations between words across lists and, more importantly, what their reaction to having noticed these relations had been. After this brief interview all subjects were thanked for their participation and fully debriefed.

Results and Discussion

Initial Tests. For each subject, we calculated a composite initial test score based on their combined performance across all five initial tests. Means for the three 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) x 5 (test) repeated measures analysis of variance (ANOVA) was applied to the initial test data. The analysis revealed no significant main effects or interaction,

indicating that all groups performed comparably across all five initial tests. In fact, all groups initially recalled approximately 50% of each study list.

Regarding our hypothesis that subjects in the two unaware groups should not be attempting to retain previously-tested materials as they proceed through the study, it is somewhat 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, if they were keeping previously tested information in mind for the final test. However, the data do not show any evidence that this occurred. The design of the present study provides a clue as to the nature of this outcome. Specifically, the related nature of the words, both within and between lists, may have boosted performance for all groups so as to obviate any effect of whether or not information was being held in mind (e.g., Horton & Petruk, 1980; Woodward & Bjork, 1971). Regardless, this null outcome on the initial tests does not preclude our finding interesting differences among these groups in final recall performance.

Final Test. Mean final recall performance is also presented in Figure 1 (right panel). A one-way ANOVA revealed a significant main effect of condition, $F(3, 156) = 29.24$, $MSe = .014$, $p < .001$. 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 the untested group, $t(78) = 9.53$, 6.98 , and 5.91 , respectively, $ps < .001$. In addition, the aware group recalled significantly more words on the final test than both the unaware group, $t(78) = 2.50$, $p = .015$, and the unaware-cue group, $t(78) = 2.26$, $p = .026$. The two unaware groups, in turn, 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 the final test. The remainder of our analyses focus on these two particular patterns of results and serve to elucidate the nature of their origins.

Testing. All three groups receiving initial testing recalled significantly more words on the final test than 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, in press). To gain further insight into the nature of this outcome, words recalled on the final test were partitioned into their origins of study. That is, we were now 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, an analysis motivated by a previous study (Tulving & Thornton, 1959).

Figure 2 illustrates the partitioning of the final free recall data by 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, with recall better for lists four and five ($M = .51$) than for lists one and two ($M = .42$). In contrast, the untested group exhibited an effect of primacy, with recall better for the first list (.37) relative to later lists ($M = .25$). Confirming this apparent relationship between testing and resulting pattern of recall, a 2 (group) x 5 (list) repeated measures ANOVA, revealed a significant group by list interaction, $F(4, 632) = 14.43$, $MSe = .019$, $p < .001$. As a quick check to verify that the observed effects of recency and primacy were indeed real, we compared mean list 1 recall (in the final test)

to mean list 5 recall (in the final test) for each group. The tested group, as a whole, showed significantly better recall of the more recently learned list words, $t(119) = -5.15$, $p < .001$, while 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 the figure suggested that the testing effect in final recall was only apparent after list one. Statistical tests confirmed this observation showing that the initially tested subjects recalled reliably more words than the initially untested subjects for study lists two, three, four, and five, $t_s(158) = 5.93, 7.34, 8.52, \text{ and } 7.93$, respectively, $p_s < .001$, but not for list one ($p > .20$).

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 both of the unaware groups. In order 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 receiving both initial and final testing. Such an analysis is typically concerned with conditionalizing recall performance in one trial (recalled/not) based on recall performance from a previous trial (recalled/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 been previously 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 the obtained difference between the aware group and the two unaware groups.

Figure 3 compares inter-test forgetting (CN) and reminiscence (NC) across the three groups of interest, the remaining two conditional probabilities being complementary. The aware group was *less* likely to forget words that had been initially recalled (CN) than the unaware group, $t(78) = 3.26, p = .002$, and the unaware-cue group, $t(78) = 3.79, p < .001$. In addition, the aware group was *more* likely to recover initially missed words (NC) than 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 one another in either respect. Considering the comparable performance of the three groups in initial testing, these novel results appear to support our hypothesis that the expectation of a final test may enhance long-term retention.

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

Subjects were allotted eight minutes for the final recall test. They were instructed to draw a line every minute under the last word they had recalled, allowing for the construction of 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 one minute of recall, at which point the number of words recalled by the aware, unaware, and unaware-cue groups had already reliably exceeded that recalled by the untested group, $t(78) = 6.25, 4.99, \text{ and } 5.13$, respectively, $ps < .001$. The effect of expectancy did not manifest itself until the second minute of recall, at which point the number of words recalled by the aware group had 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 serve to further 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. First, taking an initial test enhanced performance on the final test (relative to no initial test). Second, having an expectation of the final test enhanced final test performance, relative to not having that expectation. Third, an explicit cue to forget was shown to be unnecessary in producing the observed difference between subjects expecting the final test and those not expecting that test. Finally, conditional analyses revealed that holding an expectation of the final test kept list words in a more accessible state, reducing forgetting between an initial and final test 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 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 to 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 (i.e., 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 explained as being due to the influence of learning of later lists interfering with retention of earlier lists (retroactive interference), at the time of final recall (Tulving & Thornton, 1959). Just as in the present study, Tulving and Thornton observed no difference in the recall of study lists during initial testing, with a marked effect of recency emerging in final free recall. In a thorough investigation of the phenomenon, manipulating the degree to which study lists had been subjected to proactive and/or retroactive interference, they found no evidence that proactive interference affected final recall performance, but they did find a significant influence of retroactive interference (see also Tulving & Psozka, 1971; for an alternative explanation see Sahakyan, Delaney, & Kelley, 2004).

As for the untested group, a manipulation not examined by Tulving and Thornton (1959), this is the first evidence for primacy in recall in the context of a multi-list learning experiment (although 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 to occur as intrusions in the recall protocols of subsequently tested lists than were words from lists that had been tested earlier. This observation, along with the finding that tested groups show 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 might help explain why the untested group displayed such poor retention of materials following the first list, while the tested groups showed initially proportionate learning of all lists.

The finding of particular interest in the present report was that, given initial testing, holding an expectation of a final test led to better performance on that test, as compared to not having that expectation. A previous study (Shimizu, 1996; see also Runquist, 1983) using unrelated word lists also hinted at this effect. Our using words related both within and between lists facilitated additional processing of previously-tested materials; this feature of our design explains why our pattern was more pronounced. In addition, the two groups not expecting the final test were shown not to differ from one another, suggesting that an explicit instruction to forget was not necessary 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.

Interestingly, 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 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 the 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 might have been more likely to notice the relations of words across lists, in effect setting up strong associations among all study materials (see Masson & McDaniel, 1981). Interrelating of list items could easily have occurred due to the overlapping of categories across lists (Rundus, 1971, Experiment 4). In contrast, a subject not expecting the final test would have no incentive to pay particular attention to such relations across lists and should, therefore, benefit less from such occurrences. Subject reports, across all conditions, suggested that previous words were in fact coming to mind during study of subsequent lists. In line with our interpretation, subjects expecting the final test consistently reported taking particular 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, then 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 (e.g., Bousfield & Bousfield, 1966; for a review see

Shuell, 1969), indicating that the relations of words across lists influenced later output. Given that subjects in the aware group had reported paying particular attention to these relations, whereas 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 that test. An examination of the final recall protocols (see Table 1), using the adjusted-ratio-of-clustering (ARC) 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 the unaware-cue group, $t(78) = 3.70, p < .001$. Again, the two unaware groups did not differ from one another, $t(78) < 1$. These data nicely complement subjective reports. Anticipation of the final test led to greater benefits procured from the relations of words across lists, probably because holding an expectation of the final test encouraged relation of 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 Jacoby and Bartz (1972), who, in examining the role of rehearsal in transfer to long-term memory, 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. In their study, terminal list words were better recalled during initial testing following an unfilled interval or with no delay (between study and test) 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 (e.g., Craik & Lockhart, 1972; Woodward, Bjork, & Jongeward, 1973) of those words, as compared to 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 based on the more elaborate processing during study.

In the same vein, subjects expecting the final test in the present study appeared to use their knowledge that they would receive this test to guide the manner in which they processed study words (across all lists), in effect setting up relations that would benefit final recall performance. Conversely, subjects not expecting the final test had no reason to pay attention to the relations of words across lists and rather 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 was no longer relevant.

In the typical classroom setting, material is taught in the form of units, each of which are 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 (e.g., five lists of 18 words, a one-minute delay between study and initial test, a 30-minute delay between initial and final testing) are certainly not representative of those of

the classroom, it seems possible 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 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.

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, 463-471.
- 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 and 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 and 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 and 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). 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. (1969b). Poststimulus output specification and differential retrieval from short-term memory. *Journal of Experimental Psychology*, 82, 168-174.
- Epstein, W. (1970). Facilitation of retrieval resulting from post-input 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 and 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.

Lundberg, 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). New York: Elsevier, Academic Press.

Masson, M.E.J., & McDaniel, M.A. (1981). The role of organization 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. (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 (On the analysis of memory activity and the conception process). *Zeitschrift fur Psychologie*, Suppl. 5.

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.

Roediger, H.L., & Karpicke, J.D. (in press). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*.

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.

Sahakyan, L., Delaney, P.F., & Kelley, C.M. (2004). Self-evaluation as a moderating factor in 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.

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 and 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.

Author Notes

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 has been presented as a poster at the 17th annual convention of the American Psychological Society in Los Angeles, CA. This manuscript benefited greatly from comments provided by Jason Chan, Sean Kang, Jessica Logan, and Mark McDaniel. Correspondence concerning this manuscript can be sent to the first author at karl.szpunar@wustl.edu or at the Department of Psychology, Campus Box 1125, Washington University in St. Louis, One Brookings Drive, St. Louis, MO 63130.

Footnote

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 at a later time (e.g., Meyer, 1934, 1936, for a review see Lundeberg & Fox, 1991), a phenomenon different from the one of interest in the present report.

Table 1.

The mean adjusted-ratio-of-clustering (ARC) scores and standard deviations for groups receiving both initial and final testing.

<u>Condition</u>	<u>ARC</u>	
	<u>M</u>	<u>SD</u>
Aware	0.66	0.16
Unaware	0.51	0.17
Unaware-Cue	0.52	0.23

Note. Range for ARC = 0-1, where 0 represents no clustering and 1 represents perfect clustering.

Figure Captions

Figure 1. The mean proportion 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. Final recall performance as a function of original study list and condition. Error bars display standard errors of the mean.

Figure 3. The 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.

Figure 1.

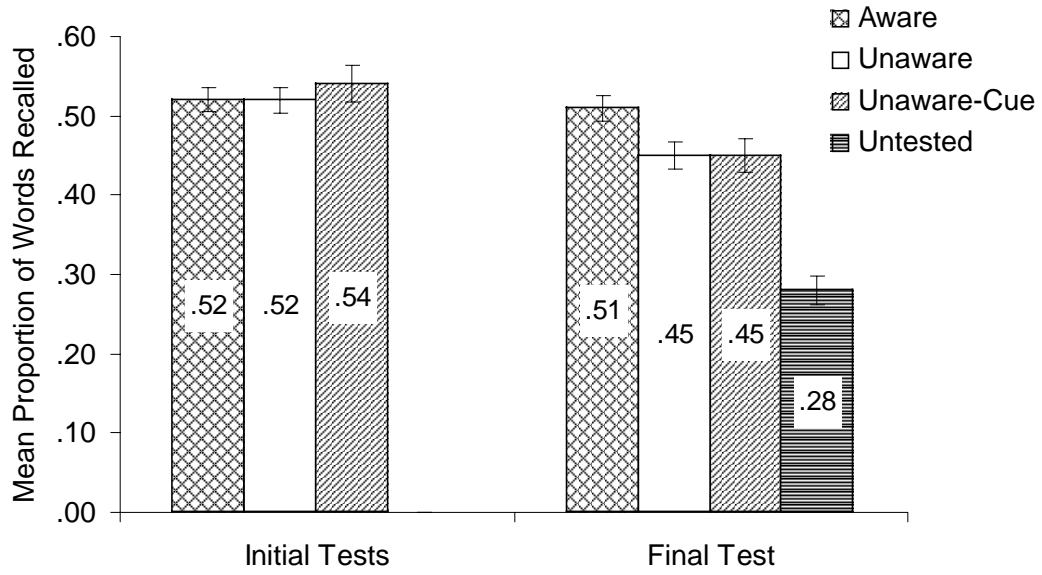


Figure 2.

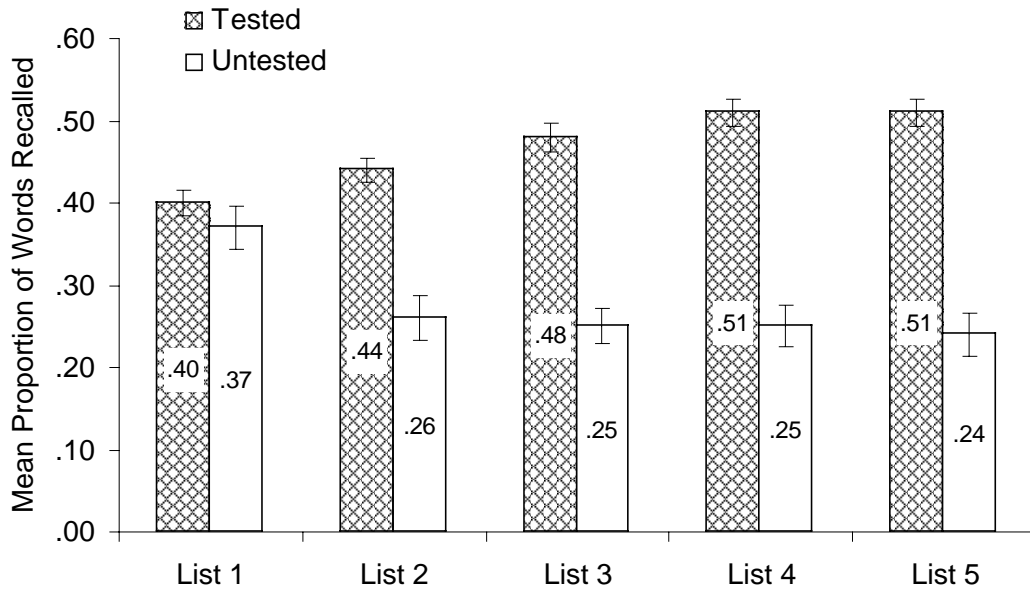


Figure 3.

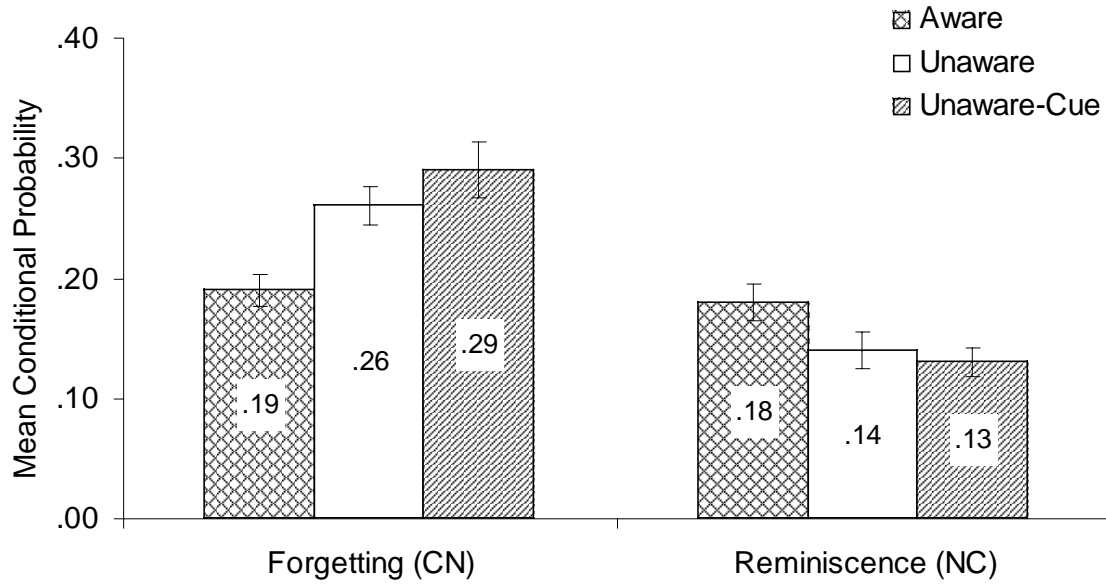


Figure 4.

