March, 2007

The Testing Effect in Recognition Memory: A Dual Process Account

Jason C.K. Chan, *Washington University in St Louis*
Kathleen B. McDermott, *Washington University in St Louis*

The Testing Effect in Recognition Memory: A Dual Process Account

Jason C.K. Chan

Kathleen B. McDermott

Washington University in St. Louis

(Word Count: 4986)

Correspondence to:
Jason Chan
Department of Psychology
Washington University – Box 1125
St. Louis MO 63130-4899
USA

Phone: 314-935-8892
Email: *ckchan@wustl.edu*

This is a manuscript of an article from *Journal of Experimental Psychology: Learning, Memory and Cognition* 33 (2007): 431, doi:10.1037/0278-7393.33.2.431. Posted with permission. This article may not exactly replicate the final version published in the APA journal. It is not the copy of record.

Abstract

The testing effect, or the finding that taking an initial test improves subsequent memory performance, is a robust and reliable phenomenon – as long as the final test is recall. Few studies have examined the effects of taking an initial recall test on final recognition performance, and results from these studies are equivocal. In four experiments, we attempt to demonstrate that initial testing can change the ways in which later recognition decisions are executed even when no difference can be detected in the recognition hit rates. Specifically, initial testing was shown to enhance later recollection but leave familiarity unchanged. This conclusion emerged from three dependent measures: source memory, exclusion performance, and remember/know judgments.

(Word Count: 117)

The Testing Effect in Recognition Memory: A Dual Process Account

The testing effect refers to the finding that taking an initial memory test after an encoding episode enhances performance on a later memory test (relative to the case in which no initial test is taken). The effect occurs across a wide range of stimuli and is very robust (Carrier & Pashler, 1992; Chan, McDermott, & Roediger, 2006; Roediger & Karpicke, 2006) -- as long as the final criterial test involves recall. Given the sizable literature on the testing effect, it is curious that very few studies have employed recognition as the criterial measure. As a result, our understanding of the testing effect is based primarily on recall.

Why is there a dearth of reported studies using recognition? One possibility is that studies that have used recognition as the final test have produced inconclusive evidence in support of the testing effect. Indeed, some of the studies using recognition as the final criterial measure have shown benefits from the intervening recall test (Hanawalt & Tarr, 1961; Hicks & Starns, 2004; Read, 1979; Roediger & McDermott, 1995; Verde, 2004) and others have not (Darley & Murdock, 1971; Jones & Roediger, 1995; Lockhart, 1975). In the current report, we consider the possibility that although testing may not consistently alter the probability of final recognition, it can change the underlying processes with which recognition decisions are executed. In light of this possibility, we examined whether an application of the dual-process framework (Jacoby, 1991) can reconcile the conflicting findings.

For decades, dual-process frameworks have been critical to our understanding of recognition memory (Yonelinas, 2002); this framework suggests that recognition performance is determined by the relative contribution of a recollection process that is under conscious control and a familiarity process that is more automatic. Recall is thought to rely heavily on recollection whereas recognition depends more evenly on recollection and familiarity. We can tentatively apply this logic to the testing effect as follows. The testing effect is consistently found in recall because taking a prior recall test has a profound effect on later recollection. When a final recognition test is administered, however, such benefits are masked by the greater contribution of familiarity. In fact, it is possible that testing could enhance recollection without affecting the recognition hit rate if participants in the no-testing condition could compensate for their lack of recollection by relying more heavily on familiarity.

Given the explanatory power of dual process frameworks, it is curious that they have yet to be formally applied to the understanding of the testing effect (see, Roediger & Karpicke, 2006, for a review). In three experiments, we attempted to reveal the effects of initial testing on the relative contribution of recollection and familiarity in a subsequent recognition test. A modified source discrimination procedure (Johnson, Hashtroudi, & Lindsay, 1993), the opposition procedure (Jacoby, 1991), and the remember/know procedure (Tulving, 1985) were used to provide converging evidence for the hypothesis that prior testing can affect the way later recognition decisions are executed even when overt changes are not detected in the hit rates. Specifically, we predicted that initial testing could enhance recollection even without a concomitant increase in overall recognition probabilities.

Overall Design

Figure 1 displays a schematic of the experimental design of Experiments 1a and 1b. Specifically, participants either studied lists of words under the *testing* condition, in which they took an immediate free recall test after the encoding of each word list, or under the *no-testing* condition, in which they performed an arithmetic task in place of the immediate recall test. Participants took a final recognition test following the presentation of all word lists and the immediate recall/arithmetic tasks. Experiments 1a and 1b used a modified source recognition test and the opposition procedure, respectively, to assess the effects of testing on later recollection and familiarity. Experiment 2 used the remember/know procedure to assess the phenomenological experience accompanying recognition decisions for items in the testing and no-testing conditions. Experiment 3 was designed to address possible criterion shift issues that might have influenced the results of earlier experiments.

Experiment 1a & 1b

With the exception of test instructions, the materials and procedures were the same for Experiments 1a and 1b; therefore, the methods for these experiments are presented together.

Method

Participants. Forty-eight undergraduate students at Washington University participated in return for course research credit (24 participants in each experiment). Participants were tested individually, in pairs, or in groups of three.

Materials. One hundred sixty unrelated words were selected (Kucera & Francis, 1967) on the basis of their length (3-10 letters), part of speech (adjective, verb, and noun), and frequency (5 – 200, $M = 64.96$). The words were broken down into eight groups of 20 words each for counterbalancing purposes, and the average word frequency in each group was roughly equal (64.15 – 66.55). For each participant, two groups of words were assigned to be studied words in the testing condition, two groups were assigned to be studied words in the no-testing condition, and the remaining four groups were assigned to be nonstudied lures in the two recognition tests. Assignment to testing/no-testing and studied/lure conditions was counterbalanced across participants.

Procedure. During the study phase, participants studied a list of words (which took a total of 90s) and then took a free recall test (90s, written recall) in the testing condition. They then performed an arithmetic task (90s), studied a second list of words (90s) and then performed another free recall test (90s) on that list. Finally, participants took a recognition test that contained words from both studied lists along with lure words.

In the no-testing condition, participants studied a list of unrelated words (90s) and performed an arithmetic task (180s). They then studied another list of words (90s), performed another arithmetic task (90s), and then took a recognition test similar to the one in the testing condition. The order in which participants completed the testing and the no-testing condition was counterbalanced. The intentional learning instructions informed participants that each word would be presented for 4s (with a 500ms inter-stimulus-interval, ISI) and that a free recall or an arithmetic task would follow the presentation of each word list. During each 90s recall test, participants recalled words from the immediately preceding word list. During the arithmetic task, participants performed mental calculations and then typed in their answers. The math problems (e.g., $42 + 27 - 3 = ?$) were shown on the screen one at a time for 10s each. After presentation of the second study list, participants performed another 90s recall test in the testing condition and a 90s arithmetic task in the no-testing condition.

In Experiment 1a, participants were asked to make one of four responses to each word in the 80-trial *modified source recognition test* (adopted from Drosopoulos, Wagner, & Born, 2004). In this test, participants were told to (a) press the "1" key if a word had appeared in the first study list, (b) press the "2" key if a word had appeared in the second study list, (c) press the "f" key if a word had been studied but they could not remember the list membership, or (d) press the "n" key for a nonstudied word. Responses were classified into five categories: (a) correct source judgments occurred when participants correctly identified the list membership of a studied item (either list 1 or list 2), (b) incorrect source judgments occurred when participants incorrectly claimed a list 1 item as coming from list 2 and vice versa, (c) no-source judgments occurred when participants pressed the "f" key for a studied item regardless of list membership, (d) correct rejections occurred when participants pressed the "n" key for a nonstudied item, and (e) the hit rate was calculated by subtracting the miss rate (claiming a studied word "new") from 1, which was equivalent to the sum of the probabilities of correct source judgments, incorrect source judgments, and no-source judgments.

The recognition test in Experiment 1b was an exclusion test, in which participants were asked to identify only items from list 2 as "old". No standalone *inclusion test* was administered. Instead, the hit rate from list 2 served as the *inclusion score* and the false alarm rate from list 1 served as the exclusion score. This procedure was adopted from Gruppuso and colleagues (Gruppuso, Lindsay, & Kelley, 1997; and see Jacoby, Shimizu, Velanova, & Rhodes, 2005, for a discussion of the potential problems with using the traditional inclusion-exclusion procedure). Specifically, participants were told to press the "o" key (for old) if they identified a word as coming from list 2. For all other words, which included words from list 1 and nonstudied words, they were to press the “n” key (for new). If participants correctly pressed “o” for a list 2 word, it was scored as a hit; if participants incorrectly pressed “o” for a list 1 word, it was scored as a false alarm.

Results for Experiment 1a

Alpha level was set at $p = .05$. Partial eta squared (*pes*) indicates effect size for analyses of variance (ANOVAs) and Cohen’s *d* indicates effect size for t-tests. An examination of the top half of Table 1 reveals that taking an initial free recall test had little impact on the overall recognition probabilities. Specifically, neither the hit rates (.88 for testing and .85 for no-testing), $t(23) = 1.44$, $p > .05$, nor the correct rejection rates (.86 for testing and .83 for no-testing), $t(23) = 1.42$, $p > .05$, differed between the testing and the no-testing conditions. However, the results from source judgments revealed that initial testing did exert an influence on the ways later recognition decisions were executed. Specifically, for studied items, initial testing increased the probability of correct source judgments (.65 for testing and .54 for no-testing), $t(23) = 2.51$, $d = .58$, and reduced the probability of incorrect source judgments (.09 for testing and .13 for no-testing), $t(23) = 2.72$, $d = .61$, but did not affect the probability of no-source judgments (.15 for testing and .19 for no-testing), $t(23) = 1.23$, $p > .10$. In sum, source memory was enhanced via initial testing despite no significant testing effect in the hit rate.[1]

Results for Experiment 1b

The bottom half of Table 1 displays the results for Experiment 1b. Interestingly, unlike Experiment 1a, there was a significant (6%) testing effect on the recognition hit rate (for list 2 items), $t(23) = 2.21$, $d = .44$, and the correct rejection rate for nonstudied items (3%), $t(23) = 2.72$, $d = .39$. More importantly, there was a large difference in the false alarm rate for list 1 items between the testing and no-testing conditions. Specifically, participants correctly rejected 77% of the list 1 items in the testing condition, but they could only reject 52% of those items in the no-testing condition. When the hit rate of list 2 (inclusion) and the false alarm rate of list 1 (exclusion) were entered into process dissociation equations, the resulting data revealed a significant interaction between testing condition (testing, no-testing) and retrieval processes (recollection, familiarity), $F(1, 23) = 17.73$, *pes* = .44 (see bottom half of Table 1). Specifically, testing doubled the probability of recollection in this experiment (.60 for testing and .30 for no-testing), $t(23) = 4.91$, $d = 1.32$, but reduced the probability of familiarity (.57 for testing and .69 for no-testing), $t(23) = 2.19$, $d = .51$.

How could such a great increase in recollection be hidden in the hit rate? One possibility is that participants relied more on familiarity during the recognition test that was not preceded by an initial test than during the recognition test that was preceded by an initial test. An increased reliance on familiarity in the no-testing condition might have arisen from subjects' realization that their source memory was impoverished; the outcome would be similar to a criterion shift. Such a retrieval strategy would lead to a lower estimate of recollection and a higher estimate of familiarity for the no-testing condition compared with the testing condition. We return to a more detailed consideration of this issue following the presentation of Experiment 2.[2]

Experiment 2

Several previous studies have investigated the effect of taking an initial recall test on the subjective experience of later remembering and knowing. The main conclusion from these experiments is that initial recall increases remember, but not know, judgments. However, individual experiments differed in their approach and the data were not always consistent. For example, Jones and Roediger (1995) and Roediger and McDermott (1995) found that taking an initial free recall test enhanced later remember responses in a final free choice recognition test (relative to no initial test). McDermott (2006) and Verde (2004) also found a similar pattern using a final free recall test and a paired-associate recognition test, respectively. However, when hit rates were taken into account by using conditional instead of raw remember probabilities (see Chan & McDermott, 2006; Rajaram, 1993), the latter studies (McDermott, 2006; Verde, 2004) did not show any influence of prior testing on remember judgments (conditional probabilities of remember/know are obtained by dividing the raw probabilities of remembering and knowing by the hit rate; therefore, if the hit rate is 80% and the raw probability of remember is 40%, the conditional probability of remember is 50%).

In the current experiment, we sought to provide further support for the hypothesis that the testing effect can be revealed in the recollective component of recognition using remember/know judgments. The general design of this experiment was the same as Experiments 1a and b except that participants studied more words (120), that a 15 min distracter task was inserted immediately prior to the recognition test, and that a single recognition test included words in both the testing and the no-testing conditions.[3]

Method

Participants. Twenty-four undergraduate students participated.

Materials. In addition to the 160 words used in Experiments 1a and b, 80 extra words were chosen using the same selection criteria. Inclusion of these new words did not change the overall averaged word frequency from that of Experiments 1a and b (M = 65.01).

Procedure. Participants studied three lists in the testing condition and three lists in the no-testing condition. The immediate free recall and mental arithmetic procedures were the same as in previous studies. In the previous experiments, word lists in the same testing condition always appeared in the same test block for the purpose of the source/exclusion test; no such restriction was present in the current experiment because only one recognition test was administered. Instructions for the pre-recognition phases were the same as in previous experiments. Prior to the recognition test, participants performed a distracter task (i.e., played Tetris) for 15 minutes. The recognition test instructions asked participants to make one of three judgments (remember, know, or new) for each word in the recognition test. Specifically, participants were told to press the “R” key (for Remember) if they could recollect specific details associated with a word’s presentation during the encoding episode, the “K” key (for Know) if a word was familiar but that they could not remember any specific details associated with it, and the “N” key for a nonstudied word.

Results and Discussion

As can be seen in Table 2, taking an initial free recall test did not change the overall hit rate in the final recognition test (.79 for testing and .76 for no-testing), $t(23)$ = 1.45, $p > .10$. Nonetheless, initial testing increased the probability of remember responses relative to the no-testing condition, and this pattern holds whether one considers the raw probability of remembering (.52 for testing and .44 for no-testing), $t(23) = 3.02$, $d = .38$, or the conditional probability of remembering (.66 for testing and

.57 for no-testing), $t(23) = 3.24$, $d = .42$. In contrast, taking an initial test reduced the raw (.26 for testing and .32 for no-testing), $t(23) = 2.50$, $d = .38$, and conditional (.34 for testing and .43 for no-testing), $t(23) = 3.24$, $d = .42$, probabilities of knowing. However, because a remember response preempts a know response, the independent remember/know (IRK) conversion was used to estimate the contribution of familiarity (see, Jacoby, Yonelinas, & Jennings, 1997, for details). We chose this method because it is arguably the least controversial method in converting remember/know data into estimates of recollection and familiarity (Dunn, 2004). Using the familiarity estimates provided by this conversion, it was clear that taking an initial recall test did not affect the contribution of familiarity in the final recognition test (.54 for testing and .57 for no-testing), $t(23) = 1.40$, $p > .10$.

Experiment 3

Data from the previous experiments show converging evidence for the hypothesis that taking an initial recall test enhances later recollection even without a comparable enhancement in the hit rate. We have mentioned that this can happen if participants used different response criteria depending on whether or not the recognition test followed a recall test. Specifically, when recollection for items on a recognition test is weak (e.g., the no-testing condition), participants may adopt a more lenient response criterion, which would be revealed as an increase in familiarity in the no-testing condition compared with the testing condition. Experiment 3 was designed to minimize the impact of such a criterion shift on the estimates of recollection and familiarity.

In this experiment, we included words from both the testing and the no-testing conditions in a single recognition test similar to that employed in Experiment 2, because participants typically do not switch their response criterion on a trial-by-trial basis (Wixted & Stretch, 2000). To employ such a procedure in a process estimation design, cued-recall (instead of free recall) was used as the initial test. Specifically, participants studied two lists of unrelated words and performed a cued-recall test on half of the words after the encoding of each list; afterwards, a single exclusion test was administered. The exclusion instructions remained the same as in Experiment 1b, in which participants were instructed to claim only list 2 words "old"; however, both list 1 and list 2 words now included tested and nontested words.

Method

Participants. Twenty-four undergraduate students participated either for research credit or for $5.

Materials. Due to the nature of the word-stem cued recall test, a new set of 120 words was chosen to ensure that none share the same first three letters. The 120 words were separated into eight sets of 15 words each for counterbalancing purposes. The mean frequency of each set ranged from 34.33 to 35.00.

Procedure. Participants studied two 30-word lists for 5s each (ISI was 1s). After presentation of each list, participants completed a 7s per trial three-letter word stem cued-recall test on half (15) the items on that list. For example, if the word stem was "but______" and one of the studied words was "butter", participants would type in "butter". Studied items presented during the cued-recall phase, regardless of whether they were correctly recalled and regardless of list membership, were assigned to the testing condition. Conversely, studied items not presented during the cued-recall phase belonged to the no-testing condition. Assignment of words to the testing and no-testing conditions was counterbalanced across participants. After the cued-recall phase for the second list, participants completed the exclusion test. The instructions for this recognition test were the same as those in Experiment 1b.

Results and Discussion

As can be seen in Table 3, initial testing increased the probability of the List 2 hit rate by approximately 13%, $t(23) = 3.31$, $d = .83$. An ANOVA revealed a significant interaction between testing condition (testing, no-testing) and process estimates (recollection, familiarity), $F(1, 23) = 4.66$, *pes* = .17 (and the main effects were also significant, both *F*s > 9.08). A planned comparison revealed that initial testing greatly enhanced recollection of the studied words (.47 for testing and .27 for no-testing), $t(23) = 3.38$, $d = .80$. But more important for the current purpose, initial testing appears to have no effect on familiarity (.55 for testing and .59 for no-testing), $t < 1$. The fact that there was no effect of initial testing on the estimates of familiarity in this experiment suggests that having participants perform a single recognition test significantly reduced the likelihood of a criterion shift, and this manipulation has permitted the effects of initial testing to be revealed without the confounds of varying response criteria.

General Discussion

In four experiments, we have demonstrated the effects of testing on subsequent recollection of studied words even when the hit rate fails to show a benefit. As we have mentioned in the Introduction, the testing effect is not always found when the final test is recognition, but it is, to our knowledge, virtually always found when the final test is recall. Results from the current study have provided strong support for the hypothesis that this discrepancy is based on the differential involvement of recollection and familiarity between recall and recognition.

We now consider previous studies that have and have not shown the testing effect in the hit rates through the perspective of the dual process framework. The idea is that the testing effect should be revealed in the hit rate if the test requires/encourages controlled retrieval of studied items. Indeed, recognition studies that have shown a testing benefit are consistent with this prediction. An examination of the studies in which testing produced an increase in recognition performance reveals that all these studies used lures that were difficult to reject based on familiarity. For example, Hanawalt and Tarr (1961) used a three-alternative forced choice recognition test that included two lures on each trial: one of them a synonym of the target and the other an antonym. Distractors that are semantically related to targets are difficult to reject based on familiarity (Benjamin, 2001; Chan & McDermott, 2006; McDermott & Chan, 2006). Roediger and McDermott (1995) also found a testing advantage in the hit rates. In their Experiment 2, participants studied semantically related words and performed initial recall tests on half the lists. During the final recognition test, participants had to distinguish studied targets (e.g., bed, rest, etc.) from their related lures (e.g., sleep). Lures that are semantically related to studied items could not be rejected based purely on familiarity. Hicks and Starns (2004) also found a recognition advantage for studied items that have received retrieval practice. Similar to the above two studies, these researchers also used lures that were semantically related to the targets. Read (1979) used a face recognition procedure and found a benefit for the testing group in final recognition probabilities. Recollection was imperative to good performance in this study because participants were told to make a positive recognition response only to faces that were shown in the *same perspective* as when they were presented during encoding. Verde (2004) also found an advantage for initial testing on final recognition hit rates in a paired-associates recognition test (i.e., participants made recognition judgments on word pairs, not single words). Similar to Read's study (1979), the lures in Verde's study were based on words that had been presented during the encoding phase. Specifically, these lures were rearranged word pairs. For example, if the word pairs "shark spinach" and "salmon tomato" were studied, then a recognition lure could be "shark tomato". Rejection of such rearranged pairs requires recollection (Castel & Craik, 2003).

The above summarized studies have revealed a clear pattern: when recollection is required to make educated recognition decisions, the testing effect can be revealed in the hit rate. An examination of the studies that did not show a testing effect for recognition (in addition to Experiments 1a and 2 reported here) reveals that they all followed the same general design: participants studied unrelated words, performed initial recall on half of those items, and then performed a (free- or forced-choice) recognition test in which lures were unrelated to the studied words (Darley & Murdock, 1971; Jones & Roediger, 1995). In such a situation participants could make relatively accurate recognition decisions based on familiarity; therefore, the need to invoke recollective processes was probably too weak for the testing-enhanced recollection to be revealed in the hit rate. In fact, this explanation is consistent with the testing-enhanced hit rates seen in our Experiments 1b and 3, because the lures in these exclusion tests included previously studied items (list 1 items), which could only be rejected based on recollective processes.

In sum, we contend that whether initial testing enhances subsequent recognition performance depends on the nature of the recognition test. If the recognition test encourages retrieval of source specifying details, then the testing effect should be revealed in the recognition hit rate. One can encourage recollective retrieval in a recognition test by increasing the similarity between the studied items and their distractors, by giving subjects ample time to respond during the recognition test (a response deadline has been shown to reduce the contribution of recollection), or by giving participants instructions that draw attention to the use of source specifying information (such as a source monitoring task, an exclusion task, or by emphasizing the importance of accuracy over quantity), etc. In general, when it is unnecessary for participants to invoke the more effortful recollective process, the testing effect would probably not be revealed in the hit rate. Ways to reduce the usage of recollection include making targets and distractors dissimilar, encouraging participants to respond quickly, or ensuring targets are very well encoded (e.g., it is unlikely for subjects to invoke effortful retrieval of a target that has been studied 20 times).

To conclude, using three different dependent measures, we have demonstrated that taking an initial recall test greatly enhances the probability of later recognition by recollection. Moreover, this enhancement can be revealed independent of enhancement in the hit rates. Future research may focus on how taking different forms of initial test (e.g., recall vs. recognition vs. multiple-choice) influences the relative contribution of recollection and familiarity on the final test. The current results may hinge heavily on the fact that we used recall as our initial test. A more complete understanding of the interaction between different initial test formats and recollection and familiarity will surely further our understanding of the testing effect.

References

Benjamin, A. S. (2001). On the dual effects of repetition on false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 941-947.

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

Castel, A. D., & Craik, F. I. M. (2003). The effects of aging and divided attention on memory for item and associative information. *Psychology and Aging, 18*, 873-885.

Chan, J. C. K., & McDermott, K. B. (2006). Remembering pragmatic inference. *Applied Cognitive Psychology, 20*, 633-639.

Chan, J. C. K., McDermott, K. B., & Roediger, H. L. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General, in press*.

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.

Drosopoulos, S., Wagner, U., & Born, J. (2004). Sleep enhances explicit recollection in recognition memory. *Learning & Memory, 12*, 44-51.

Dunn, J. C. (2004). Remember-know: A matter of confidence. *Psychological Review, 111*, 524-542.

Gruppuso, V., Lindsay, D. S., & Kelley, C. M. (1997). The process-dissociation procedure and similarity: Defining and estimating recollection and familiarity in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 259-278.

Hanawalt, N. G., & Tarr, A. G. (1961). The effect of recall upon recognition. *Journal of Experimental Psychology, 62*, 361-367.

Hicks, J. L., & Starns, J. J. (2004). Retrieval-induced forgetting occurs in tests of item recognition. *Psychonomic Bulletin & Review, 11*, 125-130.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*, 513-541.

Jacoby, L. L., Shimizu, Y., Velanova, K., & Rhodes, M. G. (2005). Age differences in depth of retrieval: Memory for foils. *Journal of Memory and Language, 52*, 493-504.

Jacoby, L. L., Yonelinas, A. P., & Jennings, J. (1997). The relation between conscious and unconscious (automatic) influences: A declaration of independence. In J. D. Cohen & J. W. Schooler (Eds.), *Scientific approaches to consciousness* (pp. 13-47). Mahwah, NJ: Erlbaum.

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

Jones, T. C., & Roediger, H. L. (1995). The experiential basis of serial position effects. *European Journal of Cognitive Psychology, 7*(1), 65-80.

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

Lockhart, R. S. (1975). The facilitation of recognition by recall. *Journal of Verbal Learning & Verbal Behavior, 14*, 253-258.

McDermott, K. B. (2006). Paradoxical effects of testing: Repeated retrieval attempts enhance the likelihood of later accurate and false recall. *Memory & Cognition, 34*, 261-267.

McDermott, K. B., & Chan, J. C. K. (2006). Effects of repetition on memory for pragmatic inferences. *Memory & Cognition, in press*.

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition, 21*, 89-102.

Read, J. D. (1979). Rehearsal and recognition of human faces. *American Journal of Psychology, 92*, 71-85.

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

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 803-814.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26*, 1-12.

Verde, M. F. (2004). The retrieval practice effect in associative recognition. *Memory & Cognition, 32*, 1265-1272.

Wixted, J. T., & Stretch, V. (2000). The case against a criterion-shift account of false memory. *Psychological Review, 107*(2), 368-376.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*, 441-517.

Yonelinas, A. P., & Jacoby, L. L. (1996). Noncriterial recollection: Familiarity as automatic, irrelevant recollection. *Consciousness and Cognition, 5*, 131-141.

Author Notes

The experiments reported here were supported in part by a grant from the James S. McDonnell Foundation (220020041). Some of the findings from this research report have been presented as a poster at the 2005 Psychonomic Society meeting.

Correspondence for this article should be addressed to J.C.K. Chan (email: jason.ckchan@gmail.com).

Footnotes

[1]Serial position analyses were also conducted for the process estimates of Experiments 1b and Experiment 2 (there was not enough data in Experiment 3 to produce smooth serial position curves) and they indicate that initial testing enhanced recollection of items at all serial positions, not just items in the terminal serial positions (cf., Lockhart, 1975, who claimed that only items in terminal serial positions receive a boost from initial testing).

[2] Our pattern is similar to those reported by Gruppuso and colleagues (Gruppuso, Lindsay, & Kelley, 1997) and Yonelinas and Jacoby (1996), who showed that familiarity estimates tend to be higher when exclusion decisions are made more difficult (which was the case here for the no-testing condition). For example, Gruppuso and colleagues found higher estimates of familiarity when the similarity between targets and lures was increased. Likewise, Yonelinas and Jacoby found that the estimate of familiarity increased when participants were asked to discriminate targets from lures based on stimulus characteristics not attended to during encoding.

[3]These changes were made to avoid ceiling effects in the final recognition test. A pilot study indicated that when participants studied only 80 words (for the testing and no-testing conditions combined) with a minimum delay between encoding and the recognition test, the hit rates were near ceiling for both the testing (M = .89) and no-testing (M = .92) conditions.

Table 1

Probabilities of Initial Recall and Final Recognition in Experiments 1a and 1b (SD are in parentheses)

	Testing	No-testing
Experiment 1a		
Correct Rejection Rate for Nonstudied Items	.86 (.18)	.83 (.16)
Hit Rate for Studied Items*	.88 (.09)	.85 (.12)
Correct Source	.65 (.18)	.54 (.20)
Incorrect Source	.09 (.06)	.13 (.07)
No-Source	.15 (.11)	.19 (.15)
Initial Free Recall Probability	.49 (.11)	
Experiment 1b		
Correct Rejection Rate for Nonstudied Items	.96 (.06)	.93 (.09)
Hit Rate for Items Studied in List 2	.83 (.13)	.77 (.14)
False Alarm Rate for Items Studied in List 1	.23 (.19)	.48 (.22)
Initial Free Recall Probability	.56 (.15)	
Recollection**	.60 (.20)	.30 (.25)
Familiarity**	.57 (.27)	.69 (.20)

*Hit Rates for Experiment 1a were derived from the summation of correct, incorrect, and no-source judgments. **Recollection and familiarity estimates were derived from the hit rate of list 2 items and the false alarm rate of list 1 items.

Table 2

Probabilities of Initial Recall, Final Recognition, and Remember/Know Judgments in Experiment 2 (SD are in parentheses).

	Testing	No-testing
Correct Rejection Rate for Nonstudied Items	.96 (.06)*	
Hit Rate for Studied Items	.79 (.13)	.76 (.14)
Remember (Raw)	.52 (.21)	.44 (.21)
Know (Raw)	.26 (.16)	.32 (.16)
Remember (Conditional)	.66 (.21)	.57 (.22)
Know (Conditional)	.34 (.21)	.43 (.22)
Initial Free Recall Probability	.52 (.13)	
Familiarity**	.54 (.23)	.57 (.22)

*The correct rejection rate does not belong to the testing condition per se because there was only one recognition test. **Familiarity estimates were derived using the Independent Remember/Know (IRK) procedure.

Table 3

Probabilities of Initial Recall and Final Recognition in Experiment 3. Participants were Told to Claim Only List 2 Items as Studied (SD are in parentheses).

	Testing	No-testing
Correct Rejection Rate for Nonstudied Items	.92 (.10)*	
Hit Rate for Items Studied in List 2	.80 (.14)	.67 (.17)
False Alarm Rate for Items Studied in List 1	.33 (.17)	.40 (.17)
Initial Cued Recall Probability	.68 (.13)	
Recollection**	.47 (.25)	.27 (.25)
Familiarity**	.59 (.25)	.55 (.16)

*The correct rejection rate does not belong to the testing condition per se because there was only a single recognition test. **Recollection and familiarity estimates were derived from the hit rate of list 2 items and the false alarm rate of list 1 items.

Figure Captions

Figure 1. A schematic of the experimental design of Experiment 1a and 1b.

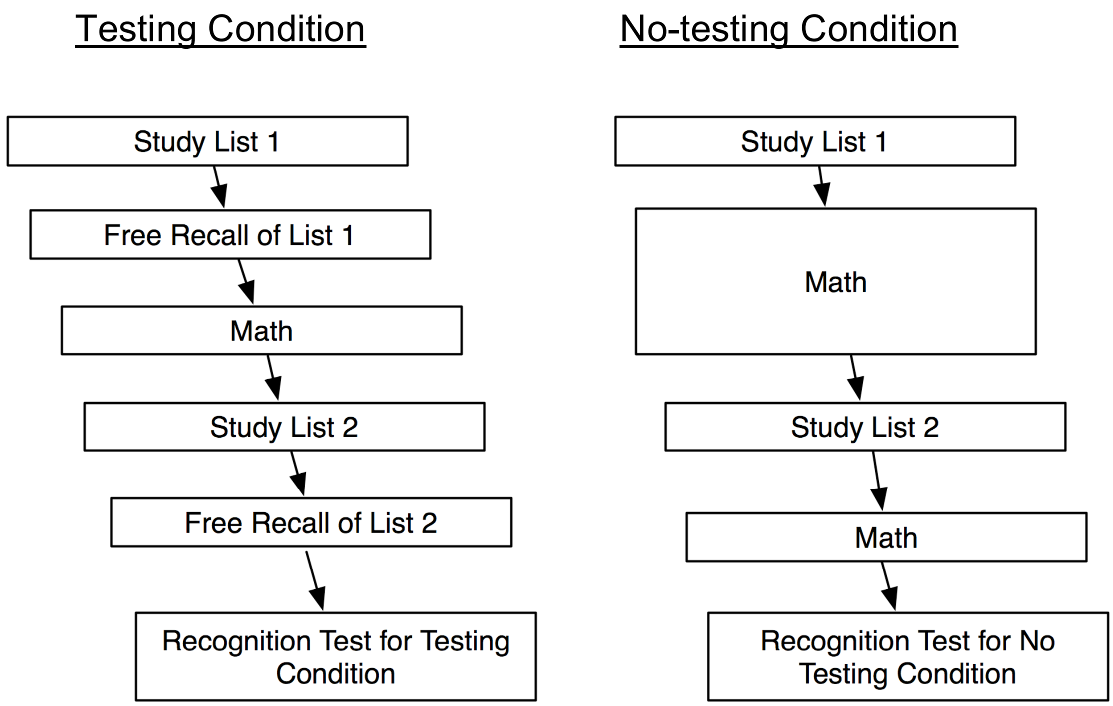
Testing Condition
Study List 1
Free Recall of List 1
Math
Study List 2
Free Recall of List 2
Recognition Test for Testing Condition
No-testing Condition
Study List 1
Math
Study List 2
Math
Recognition Test for No Testing Condition

Figure 1