

THE TESTING EFFECT AND THE COMPONENTS OF RECOGNITION MEMORY:  
WHAT EFFECTS DO TEST TYPE AND PERFORMANCE AT INTERVENING  
TEST HAVE ON FINAL RECOGNITION TESTS?

Except where reference is made to the work of others, the work described in this dissertation is my own or was done in collaboration with my advisory committee. This dissertation does not include proprietary or classified information.

---

Dale L. Smith

Certificate of Approval:

---

Ana Franco-Watkins  
Assistant Professor  
Psychology

---

Lewis M. Barker, Chair  
Professor  
Psychology

---

Adrian L. Thomas  
Associate Professor  
Psychology

---

Alejandro A. Lazarte  
Associate Professor  
Psychology

---

George T. Flowers  
Interim Dean  
Graduate School

THE TESTING EFFECT AND THE COMPONENTS OF RECOGNITION MEMORY:  
WHAT EFFECTS DO TEST TYPE AND PERFORMANCE AT INTERVENING  
TEST HAVE ON FINAL RECOGNITION TESTS?

Dale L. Smith

A Dissertation

Submitted to

the Graduate Faculty of

Auburn University

in Partial Fulfillment of the

Requirements for the

Degree of

Doctor of Philosophy

Auburn, Alabama  
August 9, 2008

THE TESTING EFFECT AND THE COMPONENTS OF RECOGNITION MEMORY:  
WHAT EFFECTS DO TEST TYPE AND PERFORMANCE AT INTERVENING  
TEST HAVE ON FINAL RECOGNITION TESTS?

Dale L. Smith

Permission is granted to Auburn University to make copies of this dissertation at its discretion, upon request of individuals or institutions and at their expense. The author reserves all publication rights.

---

Signature of Author

---

Date of Graduation

DISSERTATION ABSTRACT

THE TESTING EFFECT AND THE COMPONENTS OF RECOGNITION MEMORY:  
WHAT EFFECTS DO TEST TYPE AND PERFORMANCE AT INTERVENING  
TEST HAVE ON FINAL RECOGNITION TESTS?

Dale L Smith

Doctor of Philosophy, August 9, 2008  
(M.S., Auburn University, 2006)  
(B.S., Olivet Nazarene University, 2001)

124 Typed Pages

Directed by Lewis Barker

The testing effect has been shown to be a robust phenomenon in recall. However, there have been inconsistencies demonstrating the testing effect in recognition final tests. This has led some to suggest that recollection, but not familiarity, benefits from intervening tests. The present studies attempted to determine if differences in type of intervening test affect recognition performance at final test, as well as whether intervening tests differentially impact recollection and familiarity using the *remember-know* and source memory procedures. Results consistently demonstrated higher final test performance in intervening yes-no test conditions than in conditions that involve additional study presentations. Final performance in recall intervening test conditions was often lower than other conditions, with multiple-choice intervening test condition

typically outperforming no-test conditions, but not by a significant margin. Potential explanations for these findings include transfer appropriate processing and the effect of intervening test performance. Comparison of the final test recollection probabilities of the different intervening test conditions did not suggest an advantage for testing over additional study trials. However, additional analysis showed that a testing advantage does exist, but only for correctly recalled items. Items that were correctly identified at intervening test were also more likely to be recollected than items that were not identified at intervening test. The results demonstrate convergence between process-estimation methods and emphasize the importance of intervening test performance in the testing effect.

## ACKNOWLEDGEMENTS

I would like to first thank Lewis Barker for both his invaluable contributions to this project, and all that he has done to further my education and career. I could not have asked for a better graduate experience, and I realize that this is primarily due to his guidance and friendship. I must also thank my committee members, Ana Franco-Watkins, Alejandro Lazarte, and Adrian Thomas for their contributions. My parents, Henry and Teresa, also share much of the responsibility for any success that I attain for always knowing when to push and when to provide support, for never letting me settle for less than I could achieve, and for devoting so much of their lives to making me who I am today. Also, without Emily Stedeford's assistance at every stage of this project I would likely still be working on it today, and without her support it would certainly have been far less enjoyable. Finally, I would like to thank God, for making all of this possible.

Style manual used:

Publication Manual of the American Psychological Association (5<sup>th</sup> ed.)

Computer software used: Superlab Version 4.0®

Microsoft Word 2000®

Microsoft Excel 2000®

SPSS 16.0 ®

## TABLE OF CONTENTS

LIST OF TABLES.....	x
LIST OF FIGURES .....	xi
I. INTRODUCTION.....	1
II. EXPERIMENT 1.....	41
Method.....	41
Results and Discussion.....	47
III. EXPERIMENT 2.....	65
Method.....	65
Results and Discussion.....	66
IV. GENERAL DISCUSSION.....	73
Conclusions.....	82
REFERENCES.....	85
APPENDICES.....	102
Appendix A: Footnotes.....	103
Appendix B: Word Lists.....	106
Appendix C: Word List Means.....	110



Appendix D: Sample Instructions.....	111
Appendix E: Sample Screen Stills.....	112

## LIST OF TABLES

### TABLES

1. Experiment 1: Final Performance Rates for <i>Remember-Know</i> and Source Memory Conditions.....	49
2. Experiment 1: Final Recollection Probabilities for <i>Remember-Know</i> and Source Conditions.....	53
3. Experiment 1: Final Test Performance Comparison of Items Correctly Identified or Missed at Intervening Test.....	60
4. Experiment 2: Final Recognition Probabilities for <i>Remember-Know</i> and Source Conditions.....	67
5. Mean Correct Identification of Studied Items at Intervening Test in All Experimental Conditions.....	69
6. Experiment 2: Final Recollection Probabilities for <i>Remember-Know</i> and Source Conditions.....	72

## LIST OF FIGURES

### FIGURES

1. Types of Recognition Responding Represented along a Familiarity/Memory Strength Axis.....	7
2. Different Criterion Placements for <i>Remember</i> and <i>Know</i> Judgments According to a Signal Detection Interpretation.....	22
3. Design for Experiment 1.....	42
4. Experiment 1: Mean Intervening Test Performance by Test Type.....	48
5. Experiment 1: Comparison of Intervening Test Conditions on Final <i>Remember-Know</i> Test Performance.....	50
6. Experiment 1: Comparison of the Contribution of Hits and FA to Performance by Intervening Test Condition in Final <i>Remember-Know</i> Test Condition.....	51
7. Experiment 1: Comparison of Intervening Test Conditions on Final Test Performance in Source Memory Condition.....	52
8. Experiment 1: The Contribution of <i>Remember</i> and <i>Know</i> Responses to the Overall Final Test Hit Rates by Intervening Test Condition.....	54
9. Experiment 1: The Contribution of Correct Source, Incorrect Source, and No Source Judgments to the Overall Final Test Hit Rates by Intervening Test Condition .....	55

10. Experiment 1: Mean Conditional Probabilities of <i>Remembering</i> an Item at Final Test if the Item Was Correctly Identified or Missed at Intervening Test or Not Tested.....	57
11. Experiment 1: Mean Conditional Probabilities of Correct Source Judgment of Item at Final Test if the Item was Correctly Identified or Missed at Intervening Test or Not Tested.....	59
12. Experiment 1: Responding criterion (C) by intervening test condition .....	61
13. Experiment 1: Response Time Differences For Correctly Recollected Items, Familiar Items, and Falsely Recollected Items in <i>Remember-Know</i> and Source Memory Conditions .....	64
14. Design for Experiment 2.....	66
15. Experiment 2: Mean Intervening Test Performances by Test Type.....	68
16. Experiment 2: Overall Final Estimation Performance by Intervening Test Condition in <i>Remember-Know</i> condition.....	70
17. Experiment 2: Overall Final Estimation Performance by Intervening Test Condition in <i>Remember-Know</i> Condition.....	71

## I. INTRODUCTION

Memory has traditionally been viewed as involving an ability to retrieve past events. Although the components of the conscious experience of memory have been debated throughout history, the emergence of psychology as a scientific discipline over the past one hundred years has led to an influx of research and theories about what the memory process entails. Early in psychology's history, James (1890) attempted to describe memory, proposing that to be a memory "I must think that I directly experienced its occurrence" (p. 612). Around the same time, Ebbinghaus began the first series of systematic experiments on memory, in which he attempted to memorize sets of nonsense syllables, and recall them after varying lengths of time. Ebbinghaus (1885) further differentiated types of conscious experience associated with memories, stating that while some memories are the result of a voluntary "exertion of the will" (p.1) to call a memory back into conscious awareness, others are reproduced involuntarily, as "this accompanying consciousness is lacking, and we know only indirectly that the "now" must be identical to the "then"" (p.2). The content of the underlying processes that result in these conscious experiences and how these processes relate to the way organisms respond to previously presented stimuli continues to be examined.

Two ways to measure responding based on memory of prior experience are through recall and recognition tests. Because recognition performance is typically higher

than recall performance, recognition was initially deemed the more sensitive, or easier, measure (e.g., McDougall, 1904). Until the 1970s, tests of recall, in which a person is asked to explicitly provide information about a study episode, and recognition, which involves presenting items from a study episode and inquiring whether or not these items were previously presented, were understood as involving the same underlying construct (Lockhart, 2000). The results of several studies, particularly demonstrations of recognition failure of recallable words, directly challenged this view (Tulving & Thompson, 1973; Watkins & Tulving, 1975; Tulving, 1983). Memory models attempting to account for these and other findings in their descriptions of the relationship between recognition and recall followed. As a result, the experimental study of recognition memory flourished.

Often defined *as the judgment of previous occurrence*, recognition memory has undergone numerous changes over the past 40 years. Perhaps none of these changes has been more significant than the emergence of experimental research supporting the idea that two independent processes are at work in recognition decisions. This idea, which has existed in various forms for centuries, is now commonly known as the *dual processes theory of recognition memory*. Ebbinghaus, who is largely seen as the founder of experimental memory research, alluded to this distinction in describing his conscious experience as he examined items he had studied the previous day:

One factor in the regular course of the results obtained seems to deserve special attention. In ordinary life it is of the greatest importance, as far as the form which memory assumes is concerned, whether the reproductions occur with

accompanying recollection or not, -- *i.e.*, whether the recurring ideas simply return or whether a knowledge of their former existence and circumstances comes back with them. For, in this second case, they obtain a higher and special value for our practical aims and for the manifestations of higher mental life. The question now is, what connection is there between the inner life of these ideas and the complicated phenomena of recollection which sometimes do and sometimes do not accompany the appearance in consciousness of images? (1885, p.58)

The recollection process Ebbinghaus describes, which involves the ability to recall specific details of the encoding of a stimulus, can be contrasted with a familiarity process in which a stimulus is accompanied with a feeling or sense of having previously been exposed to it. That recognition memory consists of these separate processes of recollection and familiarity seems largely a point of agreement. Whether these two processes can be experimentally separated, however, continues to be an area of debate.

#### *Examining recognition memory processes*

The methodology now commonly used in recognition experiments was introduced by Strong (1912). He presented full page advertisements sequentially and later tested for recognition of the items by presenting the previously seen advertisements in combination with new advertisements that were not previously presented. Similar single-item recognition experiments, in which items are presented individually at study, were sparse until the 1960s. At this point, the zeitgeist of emerging cognitive psychology models led to a reemergence of the single-item methodology for studying recognition memory. This reemergence often involved presenting numerous items, such as words, individually at

study, and asking participants at test to report whether the item was previously presented (a *yes* judgment), or not (a *no* judgment). Early models treated recognition and recall as a single process, often the result of *tagging* an item at study and later examining memory in search of a tagged item (e.g., Yntema & Trask, 1963). Other single process models based on strength theory (e.g., Wickelgreen & Norman, 1966) relied on encoding information such as recency of presentation to impact a familiarity judgment.

These single process theories, however, had difficulty explaining some memory findings. Most notably, the task used at encoding often leads to differences in recognition and recall performance. For example, Eagle & Leiter (1964) gave two groups of participants a list of words. They instructed one group to memorize the words for a later test. The other group was instructed to simply classify the words based upon parts of speech, with no mention of a later test. When the memory test was unexpected, recognition performance was higher than recall performance. However, when the test was expected, recall performance was higher. If recognition and recall performance is based upon a single strength dimension, different encoding tasks should not lead to divergent results on the two test types. These and other findings led to the emergence of what would later be known as *generate-recognize models* (e.g., Bahrick, 1970; Kintsch, 1970). The generate-recognize models postulated that two processes were involved in recall memory; a search process leading to retrieval of an item, and confirmation that the item was part of the study list. Many, perhaps most notably Kintsch (1970; but see also Murdock, 1968; Bower, Clark, Lesgold & Winzenz, 1969), postulated that recognition involved only an assessment of “the newness of the trace” of the item in memory and the



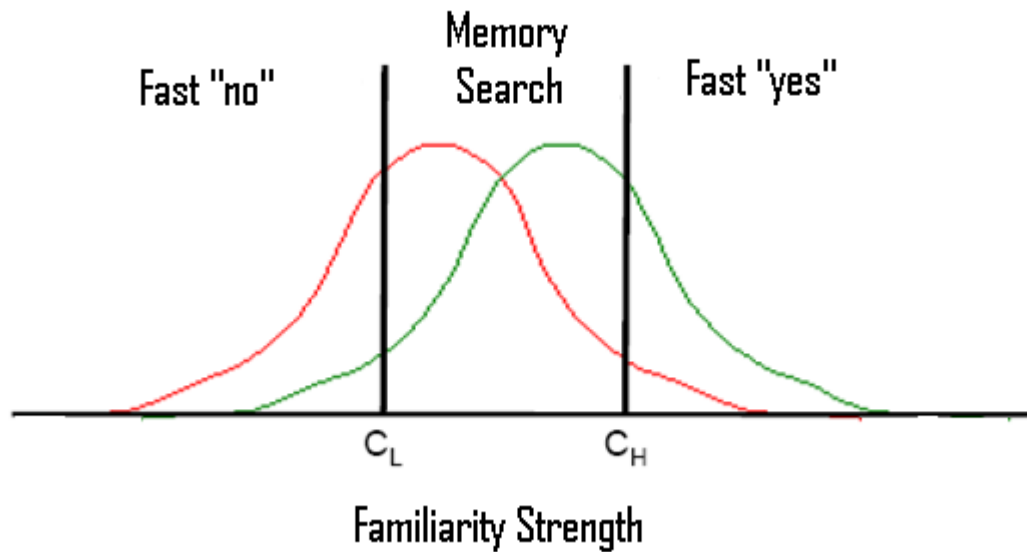
setting of a responding criterion; the search process involved in recall performance was not believed to be involved in making a recognition decision.

Several experiments provided evidence suggesting that experimental variables thought to facilitate retrieval, such as study-list organization, did not have an effect on recognition performance, but did play a large role in recall performance (e.g., Dale, 1967; Kintsch, 1968). These studies clearly supported the distinction between processes involved in recall and recognition. Few at this point had considered the idea that a search process may also be utilized in recognition. Several experiments that examined the effects of list organization on recognition later suggested this conclusion (e.g., Mandler, Pearlstone, & Koopmans, 1969; Mandler, 1972; Juola, Fischler, Wood, & Atkinson, 1971). Mandler, Pearlstone, and Koopmans (1969) used a study list organization paradigm similar to earlier research conducted by Kintsch (1968). Participants were asked to sort study words into two to seven categories. While Kintsch had found no effect of list organization on recognition performance, Mandler et al.'s study found that performance on recognition tests was reliably affected by list organization. Similar effects have been found using lists arranged hierarchically (Bower, Clark, Lesgold, & Winzenz, 1969) and syntactically (Lachman & Tuttle, 1965). Mandler (1972) proposed that two states of recognition exist. One state involves high confident items, and does not require a retrieval check. Those items that are not initially recognized with a high level of confidence, however, are subjected to a retrieval check, which should be differentially affected by changes in study list organization. Mandler & Boeck (1974) used a similar categorization procedure in which participants organized items into different numbers of

categories at study. They then assessed the effect of level of organization on recognition performance, comparing responses with long and short latencies. Their results indicated that organization had no effect on the faster responses, supporting the notion that these responses are based on an initial familiarity assessment without a search process. Organization did, however, reliably affect longer latency responses, which the authors suggested was the result of a search process.

Following a pair of memory experiments measuring response latencies, Juola, Fischler, Wood, and Atkinson (1971) independently came to a similar conclusion about the nature of recognition memory. Using well-learned words to keep the rate of incorrect responses low, they manipulated list length and distracter type and measured corresponding response latencies. Their results indicated that overall response latency increased as a function of study list size, and response latencies for *no* judgments increased when semantically or visually similar distracters were presented. Juola et al. also found that additional presentations of items at test resulted in shorter latencies to those items. The researchers interpreted these results as suggesting that participants initially made a familiarity judgment about a presented item that led to a decision about whether the item was *old* (previously presented) or *new*. In a signal detection framework memory strength is represented along a continuum. Old and new items are presented as two separate distributions that typically overlap to some extent (see Figure 1). According to Juola et al., if the familiarity value exceeded a high criterion (denoted  $C_H$ ) a fast *yes* response was given, requiring no search process. If the familiarity value fell below a low criterion (denoted  $C_L$ ) a fast *no* response was given, also absent a search process. If the

familiarity value fell between the two criteria, however, a search of the memorized list was conducted in an attempt to identify the presented item. Additional presentations of an item at test increased its familiarity value, thus leading to a faster response to that item on subsequent trials. Increases in study list size and the presence of semantically or visually similar distracters require a more exhaustive search process, and consequently, a longer latency.



*Figure 1.* . Types of Recognition Responding Represented along a Familiarity/Memory Strength Axis.

Tulving and Thompson (1971) also espoused the idea of a necessary search process in recognition, stating that theories assuming that recognition memory is automatic are misleading and “should be corrected” (p.116). They postulated that by holding encoding and storage conditions constant while manipulating conditions at test, it

could be demonstrated that accessibility of stored information differs across items. They were able to demonstrate context effects of studied words by presenting words singularly or as pairs at study, then varying single and paired items at test. Some singularly presented words and components of pairs at test were also distractor items. Their results indicated that the presentation format at test had significant effects on performance; performance was greatest in all conditions when the format at study and test was the same. Tulving and Thompson argued that such context effects would be difficult to reconcile with the idea that recognition is automatic; an automatic process should not be differentially affected by the presence or absence of other words at study or test.

Not all experimental evidence deemed the search process a necessary component of recognition memory, however. Gillund and Shiffrin (1984) compared recognition performance during two response conditions. One forced participants to make a fast (500ms) decision, and another required that subjects wait 2-3 seconds before making their decision. The assumption was that fast responding should rely on familiarity judgments (an assumption shared by earlier search process proponents), while the slower decision should rely more heavily on a participant's search process. Similar to earlier studies (Juola, et al., 1971; Mandler & Boeck, 1974), Gillund and Shiffrin found that slower responses were more accurate. However, they found no interaction between any of the variables thought to differentially affect search and familiarity process (i.e., the number of presentations of target items, study list length, depth of encoding of study items, and distracter type) and speed of responding. Gillund and Shiffrin concluded that a

search process may not be required, but can often occur if the participant chooses to initiate one.

With a few notable exceptions (e.g., Gillund & Shiffrin, 1984; Kitsch, 1968), the bulk of the experimental evidence suggested the presence of some form of search process in recognition memory, but the nature of the alleged search process varied. For example, Anderson and Bower (1974) proposed that four different types of retrieval exist in memory, not all of which were believed to be a part of recognition memory's search:

- (a) the associative chaining through long-term memory during free recall, examining idea after idea, searching for senses of words that occurred in the list;
- (b) the examination of list markers or contextual prepositions from a sense or idea in the attempt to determine whether that sense occurred in the list;
- (c) the generation of lexical realization of the sense in recall;
- (d) the access to a sense from a word. (p. 411)

Anderson and Bower postulated that while recall involves retrieval process *a* followed by *b* and *c*, the recognition search process involves *d* followed by *b*. This assertion resembles the theories discussed earlier, while clearly dissociating the search process involved in recognition with a more exhaustive search process assumed to be necessary for recall.

Mandler (1980) argued that the search process outlined by Anderson & Bower (1974) was too restrictive, and failed to take into consideration the familiarity component in recognition memory. He also noted that the exact nature of the search process depends largely on the requirements of the task. Because different tasks provide varying degrees and types of information at test, the search process required varies accordingly. For

example, being asked to recognize one component of a previously presented word pair (i.e., *train* from the pair *train-cart*) when given the corresponding word (*cart*) in a *paired associates* task will likely result in a search process that differs from that used when asked to recognize an item previously presented without such a cue.

This search process in recognition memory is now largely known as *recollection*. An assumption made in many recognition studies is that this search-based recollection process and familiarity are independent. However, not all researchers agree with this independence assumption. Curran and Hintzman (1995), for example, showed that in some cases recollection-based responses and familiarity based responses are correlated (though see Kelley & Jacoby, 2000; Jacoby, 1998 for opposing view). Several theorists (e.g., Joordens & Merikle, 1993) have also postulated that familiarity must first occur for successful recollection. This is tantamount to the generate/recognize models of recall, and eschews the notion of independence of the two processes. In contrast, Mandler's original dual process theory postulated that recollection occurs only if the familiarity process is unsuccessful. His later theories (1980), though, considered the two processes to be independent, parallel processes, an assumption inherent in most dual process theories (see Yonelinas, 2002 for review).

#### *Estimating the two processes*

Mandler postulated that results obtained via the search process involved in cued recall, using the paired associates task outlined earlier, should be a relevant indicator of recollection. He further theorized that familiarity estimates could also be obtained using the formula:

$$R_g = F + (1 - F) R \quad (1)$$

$R_g$  represents the probability of making an *old* response,  $R$  is the probability of recollection, and  $F$  is the probability of a familiarity-based response. By assuming recollection could be estimated based upon recall, Mandler derived some of the earliest quantitative estimates of recollection and familiarity.

Other researchers had different approaches to studying the underlying processes in recognition memory. Tulving's (1976) early encoding specificity theories viewed the search process as involving identification of a *trace* in memory that is established at encoding. The nature of this trace can be influenced by a number of variables, including contextual or semantic features. The function of the search process is to locate this trace in memory, a process that is influenced by the properties of the trace, as well as the information available to the individual during retrieval. Tulving (1985) later related recognition memory to two conscious processes. One consists of a conscious *autonoetic* awareness, based upon episodic memory. Tulving's episodic memory involves memory for personally experienced events. A memory accompanied by a autonoetic awareness therefore contains a conscious awareness of the personally experienced prior occurrence of an event.

The other conscious process consists of a *noetic* memory process, which is based on semantic memory, and involves awareness of prior presentation, though without the "phenomenal flavour" of episodic memory. Semantic memory, according to Tulving, is a memory store for general facts or knowledge. Tulving postulated that because *autonoetic* and *noetic* processes involve different phenomenal experiences, participants should be

able to distinguish the two consciously, and report which process influenced their recognition decision.

To test characteristics of this conscious awareness in memory, Tulving asked participants to report whether they *remembered* seeing the item on a previous list, or simply *knew* that it had been previously presented, representing auto-noetic and noetic processes, respectively. Although this *remember/know* test will be revisited in detail shortly, the notion that the search component of recognition memory contains, or is comprised mainly of, auto-noetic awareness has been largely accepted.

Certain characteristics of auto-noetic consciousness closely align it with the recollection process proposed earlier. For example, the perception of time is considered an auto-noetic process. Therefore, such awareness should be necessary for recency judgments, and likely frequency judgments. This use of conscious awareness in recognition fits quite well with Atkinson & Juola's (1974) dual process model's use of a search process following an initial familiarity assessment. Jacoby & Dallas (1981) also generally agree with the semantic/episodic distinction in recognition memory. They consider recollection a consciously controlled process, and prefer to refer more specifically to the elaboration of a word's study context as the basis of recollection. More recently, Yonelinas (2002) refers to recollection as the "retrieval process whereby "qualitative" information about a previous event is retrieved" (p. 446), a process that is based on conscious awareness.

A conscious component to recollection is largely agreed upon, but the nature of the conscious process behind the familiarity component of recognition has been more



contentious. Jacoby and colleagues (Jacoby, 1984; Jacoby, 1991; Jacoby & Dallas, 1981) view familiarity as an automatic process. This process can be based on perceptual fluency, the enhanced perceptual processing of a stimulus seen in implicit memory tasks, as well as conceptual fluency, the enhanced processing of a stimulus' meaning. Differences between performance on some implicit memory tasks and familiarity ability are attributed to the conceptual component of familiarity (Jacoby, 1984). Though Jacoby and colleagues have demonstrated that familiarity may affect perceptual fluency (e.g., Jacoby & Dallas, 1981; Jacoby & Whitehouse, 1989), familiarity and implicit memory are not the same. Some disagreement exists about whether implicit memory and familiarity are supported by the same system or rely on different memory systems, though numerous studies have demonstrated dissociability of the factors that influence familiarity and implicit memory (e.g., Light & Prull, 1995; Roediger & McDermott, 1993). The exact nature of the relationship between familiarity and implicit memory, however, is beyond the scope of this paper (for review see Yonelinas, 2002).

If we accept a definition of recollection as a retrieval process generating qualitative information about study items, a task that requires participants to produce additional qualitative information about the study event should give us some index of recollection apart from familiarity. Mandler's (1980) early estimates of recollection were derived directly from recall, and Tulving's *remember-know* procedure focused on metacognitive awareness. Several additional tasks exist which are capable of determining whether specific information about the study episode can be retrieved. One such task is the *process-dissociation* procedure.

Developed by Jacoby (1991), the *process-dissociation* procedure emphasizes the separation of recollection and familiarity based on the amount of control over responding they allow. As mentioned earlier, Jacoby and colleagues assume familiarity is an automatic process and recollection is a consciously controlled process. The *process-dissociation* procedure typically involves presenting study words in two lists or modalities, such as text and spoken words, or green and black text. Participants are then tested in two conditions. First, the *inclusion* condition simply asks if an item was presented previously, regardless of its source. The *exclusion* condition directs participants to respond only to items that were from one source and not the other. The theory behind the task is that differences in performance between inclusion and exclusion conditions should be due to familiarity decisions playing a role in inclusion condition decisions. Because both list items should be equally familiar, recollection of source must be required for exclusion condition performance. If recollection and familiarity are independent, the probability of a correct ‘yes’ response to an item from source one in the inclusion condition is:

$$P(I) = R + (1 - R)F. \quad (2)$$

The probability of an incorrect ‘yes’ response to an item from source one in an exclusion condition would then be calculated:

$$P(E) = (1 - R)F. \quad (3)$$

By subtracting equation 3 from equation 2 it should be possible to obtain an estimate of recollection as:

$$R = P(I) - P(E). \quad (4)$$

Familiarity estimates can be calculated using the estimation of recollection by:

$$F = P(E) / (1 - R). \quad (5)$$

Gruppuso, Lindsay, and Kelley (1997) modified the process-dissociation procedure to include only an exclusion task in a study that required participants to discriminate items belonging to two similar lists. Rather than use a separate inclusion task, inclusion estimates were derived from performance on excluded list items by assuming that participants respond 'yes' to items from this list based solely on familiarity. If participants recollected the source of the item as being on the excluded list it should have been rejected as instructed. Inclusion score was derived from performance on items from the included list, as such items could be reported 'old' due to familiarity or recollection of the item's source. This exclusion-only procedure was used both to simplify the process-dissociation procedure and to eliminate the possibility of a criterion shift between separate inclusion and exclusion tasks. This procedure has since been utilized in numerous studies examining recollection and familiarity under a variety of different conditions (e.g., Jones, 2006; Jones & Jacoby, 2001; Jacoby, Jones & Dolan, 1998; Chan & McDermott, 2007).

A related task involves *source memory*, or memory for details about the particular source or modality of the item at study (for review see Mitchell & Johnson, 2000). Much like the process dissociation task, a typical source memory task presents items in multiple modalities or lists. Subjects are asked at test to determine whether an item was previously presented, as well as the item's context. Items for which source memory is available should be based upon recollection, as at least some qualitative information from the

encoding episode must be available to make these judgments. Similarly, *associative recognition* tasks (Calkins, 1894) involving presentation of pairs of items at study can measure recollection. Because a word's associate at study could be considered part of its context, retrieval of this associate can be taken as a measure of recollection. Familiarity should be of limited value in the above tasks, as all items should be equally familiar.

A limitation of source memory, associative recollection and *process-dissociation* procedures in measuring recollection is that they specify the exact nature of the information required for a recollection response to be counted (inclusion in one or another modality or list or the item's associate). Recollecting additional information, such as hearing a noise in the background while encoding an item, may lead to a recollection of that item at retrieval, but may not necessarily be the exact information required to be considered a response based on recollection (i.e., color of the item or its associate). Smith, Glenberg, and Bjork (1978) demonstrated that different types of contextual factors (i.e., environmental vs. experimental) may differentially affect recall and recognition. Environmental context (i.e., birds chirping) may affect recall performance more than recognition. Given the similarities between recall and the recollection component to recognition, specifying the exact contextual elements that constitute recollection may result in recollection estimates that are artificially low.

#### *Convergence of results from process estimation methods*

Each task previously outlined is capable of deriving separate estimates of recollection and familiarity. The probing question, however, is whether convergent results between these measures can be found using manipulations assumed to

differentially affect familiarity and recollection. Such convergent results would provide further support for the existence of two processes at work in recognition decisions as well as the ability of the existing methods to assess these processes. Several studies that have directly compared source memory and *remember-know* responses have indicated that a high degree of similarity exists between the measures (but see Hicks, Marsh & Ritschel, 2002). For example, Donaldson, MacKenzie, and Underhill (1996) compared source judgments to *remember-know* judgments and found that performance on items whose source could be correctly identified was very similar to performance on items marked as *remembered*. Meiser and Sattler (2007) also demonstrated such convergence between source memory and *remember-know* methods, though they found that the type of source information retrieved may have an effect on this relationship. However, few studies have directly compared the process estimation methods. A comparison of the results from studies using these process estimation methods and the effects of factors that may differentially impact recollection and familiarity provides insight into the convergence between these methods.

Several factors have been known to differentially affect recollection and familiarity. Yonelinas (2002) reviewed all available studies using *remember-know* and process dissociation procedures, as well as the encoding and retrieval manipulations designed to affect recognition memory. At study, dividing attention, manipulating level of processing, and generating test items affect both processes, but the magnitude of these effects on the two processes differs. The first, dividing attention at study, adversely

affects both recollection and familiarity. However, Yonelinas' review of available studies suggests that the effects on recollection are typically larger than the effects on familiarity.

Levels of processing manipulations also differentially affect recollection and familiarity. Deeper, or semantic, processing leads to greater increases in recollection and familiarity than shallower, or perceptual, processing, though this effect is greater for recollection than familiarity. Yonelinas (2001) directly compared *remember-know* and *process-dissociation* estimates of recollection and familiarity for items with different levels of processing at study, and demonstrated that both processes led to similar estimates. Yonelinas (2002) reviewed 17 available studies using either *remember-know* or *process-dissociation* methods to assess the effects of levels of processing on recollection and familiarity. He found that in all but one (Java, Gregg, & Gardiner, 1997) recollection increased, and in all but three (Gardiner, 1988; Toth, 1996; Wagner, Stebbins, Masciari, Fleischman, & Gabrieli, 1998) this effect was more pronounced for recollection than familiarity.

Having to generate the list items at study, such as solving anagrams or filling in missing letters, also tends to significantly increase recollection while increasing familiarity to a lesser extent. The 11 studies reviewed by Yonelinas (2002) showed a high level of convergence between methods. Donaldson, MacKenzie, and Underhill (1996) also demonstrated that the effects of generating items at study on recollection and source memory were very similar, as would be expected if source memory relies heavily on recollection. Altogether, the results of the preceding experimental manipulations at encoding demonstrate a high level of convergence between methods, and suggest that

encoding factors play a strong role in determining whether an item will later be recognized based on familiarity or recollection.

Manipulations during test, such as manipulations of processing fluency and of the amount of time available for participants to make recognition decisions, may also differentially impact recollection and familiarity. The results of studies using process estimation methods have been less clear. Demonstrating convergence of these estimation methods are several manipulations of processing fluency, which can involve a number of procedures ranging from flashing a word prior to its presentation at test to simply presenting some words at test more clearly than others. Researchers have consistently found that such manipulations lead to an increase in familiarity-based responses but not recollection. This finding is robust across estimation methods, including associative recognition (e.g., Westerman, 2001), *remember-know* (e.g., Rajaram, 1993; but see Higham & Vokey, 2004) and *process-dissociation* procedures (e.g., LeCompte, 1995).

As mentioned above, the origins of theories involving two separate processes in recognition memory were rooted in response time studies. It stands to reason that allowing less time to respond should impair recollection to a greater degree than familiarity, though some divergence in estimation methods has been demonstrated. Studies using process dissociation have generally shown that restricting the amount of time to respond has the expected effect of decreasing recollection, but has little effect on familiarity (for review see Yonelinas, 2002). Studies of associative and source memory have also demonstrated that item recognition memory is available earlier than memory for associative (e.g., Gronlund & Ratcliff, 1989) or source information (e.g., Hintzman &

Caulton, 1997; McElree, Dolan, & Jacoby, 1999), as would be predicted if associative and source information rely primarily on a separate, more intensive, search process.

In contrast, *remember-know* studies have consistently found that *remember* responses can be made more quickly than, or as quickly as, *know* responses (e.g., Dewhurst & Conway; 1994; Dewhurst, Holmes, Brandt & Dean, 2006; Gardiner, Ramponi, Richardson-Klavehn, 1999). Yonelinas (2002) suggested that such results may be due to instructions leading to participants to make a *know* response only when an item is not recollected. However, Dewhurst et al's (2006) analysis included conditions in which participants made faster *remember* responses even though the *remember-know* procedure was separated from the old/new decision. This should have eliminated the possibility of such instructions impacting reaction times to the preceding old/new decision. These results can be easily explained by proponents of the idea that *remember* and *know* judgments do not represent qualitatively different processes. Wixted & Stretch (2004), for example, simply refer to the finding that more confident responses are typically made more quickly. Because *remember* responses are typically made with higher confidence, even *remember* false alarms should be, and are, made more quickly than *know* hits. The high degree of convergence between process estimation methods using other manipulations, and even the convergence between source memory and process-dissociation methods using reaction times, makes *remember-know* reaction time data quite puzzling.

*Theories of recognition*



Although supportive of dual process accounts, the convergence of estimation methods provides only part of the recent experimental support for dual process assertions. Numerous recent studies of the neuroanatomical substrates of recognition memory have also demonstrated the utility of such a distinction (for review see Eichenbaum, Yonelinas & Ranganath, 2007). Despite generally convergent results, additional questions remain about whether two processes are necessary to explain the aforementioned findings. A participant's ability to introspectively separate the two processes using the *remember-know* procedure in particular has been the focal point of much recent debate. If an explanation based upon a single underlying strength dimension is capable of explaining these results, it would seem to be the more parsimonious alternative. Donaldson (1996) concluded that the results of *remember-know* studies, rather than suggesting the presence of separate memory systems, are better described as demonstrating the presence of two separate response criteria. One criterion exists for *remember*, and another for *know* (see Figure 2), while the underlying distributions adhere to a signal detection framework. This signal detection framework simply involves two distributions, one for old items, and one for newly presented items, or foils, which lie along a memory strength axis. Performing a meta-analysis of available data, Donaldson contradicted the earlier claims of Gardiner and colleagues (e.g., Gardiner, 1988; Gardiner & Java, 1990) that *know* judgments could not simply be explained as weaker memories. Donaldson demonstrated that the amount of *know* responses varied according to the placement of the response criterion, and that estimates of memory were the same regardless of whether they were calculated using *remember* or *know* responses. These two findings could be easily interpreted using a

signal detection framework, but are more difficult to explain if *remember* and *know* responses reflect two separate memory components. Donaldson acknowledged that the distinction between recollection and familiarity may be a relevant one. However, he argued that the *remember-know* methodology does not capture these two processes, and is therefore not a useful measure. Hirshman and Master (1995; 1997) also modeled earlier *remember-know* experiments and came to similar conclusions regarding the ability of a single underlying signal detection process to explain the results.

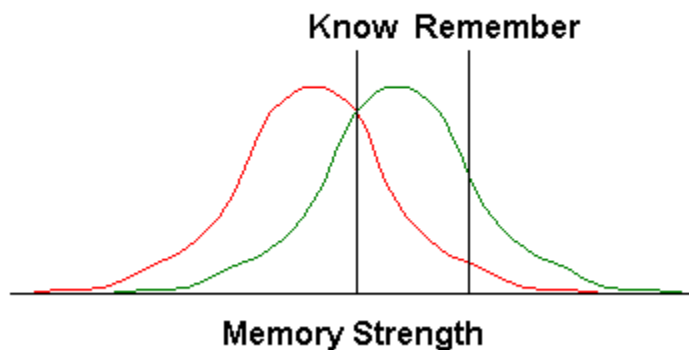


Figure 2. Different Criterion Placements for *Remember* and *Know* Judgments According to a Signal Detection Interpretation

Jacoby, Yonelinas, and Jennings (1997) suggested an error in the way *remember-know* data had been analyzed. Because participants are instructed to respond with a *remember* judgment whenever recollective detail is present, it is likely that during many of these trials familiarity is also present, but not taken into consideration. They suggested that because recollection and familiarity are independent, the probability of a *know* response when no recollective detail is present is not an accurate measure of familiarity,

and needs to be modified to take into account the presence of familiarity when a *remember* response is given. Indeed, Donaldson (1996) concluded his meta-analysis with a reanalysis of the data assuming independence between the two processes, and suggested that it led to a breakdown of the earlier stated relationships. Even though such a modification can lead the dual process account to adequately describe existing data, the unidimensional signal detection theories may still provide as good a description of these data. Dunn (2004) reviewed 72 available studies using the *remember-know* procedure, and argued that detection theory can account for all available findings. This position has recently been supported by several recent reviews that suggest that a signal detection framework better fits existing data than dual process accounts (Wixted, 2007; Rotello, Macmillan, Hicks, & Hautus, 2007, Wixted & Stretch, 2004).

At the crux of recent debate has been the use of receiver operating characteristic (ROC; Green & Swets, 1966) data to model results of remember-know experiments. Recognition memory ROC curves are typically constructed by asking participants to supply confidence ratings on their responses, and plotting hit verses false alarm rate at different levels of confidence. A signal detection framework, which assumes that memory strength exists along a continuum has historically better explained ROC data than theories that assume memory thresholds exist (e.g., Blackwell, 1953). This has led to the abandonment of most threshold-based theories over the past forty years. However, some have suggested that recent ROC data collected using remember-know or source procedures is evidence for the existence of recollection thresholds (e.g., Yonelinas, 1997; 1999; Slotnick & Dodson, 2005; Healey, Light, & Chung, 2005). These theories maintain

that familiarity still exists along a continuum best described by signal detection theory, but that recollection is fundamentally different, and operates on an ‘all-or-nothing’ threshold basis. Although a further analysis of how threshold-based and signal detection-based theories explain existing ROC data is beyond the scope of this paper<sup>1</sup> (for review see Wixted, 2007; Parks & Yonelinas, 2007), both theories can adequately explain most ROC data quite well.

*Separating the two processes consciously*

It is important to note that many signal detection theorists, most notably Wixted (2007; Wixted & Stretch, 2004), do not dispute the notion that two separate processes are involved in recognition memory. However, like Donaldson (1996), Hirshman and Master (1995; 1997), and Dunn (2004), they question the ability of participants to respond based solely on recollection or familiarity, postulating the two processes are not separable at retrieval. In fact, Wixted (2007) refers to his theory as a Dual-Process Unequal-Variance Signal Detection Theory (UVSDT), according to which, both familiarity and recollection are continuous variables that are combined to lead to a recognition decision. Most researchers accept the intuitive and experimentally supported notion that two processes are involved in recognition memory. While some minor differences of opinion about the nature of these two processes exist, most dual process theorists view the dissociation between these processes as involving conscious awareness, intentional control, or response confidence (Yonelinas, 2001). A current, and controversial, question remains as to whether these two processes can be metacognitively separated at test using the *remember-know* procedure.

Wixted & Stretch (2004) argue that *remember* judgments are simply stronger, more confident, memories, and do not differ in kind from *know* judgments. They demonstrated the effectiveness of such a detection account to accommodate some findings (such as *remember* false-alarms) that dual process theorists have typically struggled with. *Remember* false alarms are problematic for theories espousing the threshold nature of recollection, as items that were not previously studied should not be expected to exceed the high recollection threshold. Early studies typically explained such false alarms as guesses, paying little attention to the phenomenon. Wixted & Stretch (2004), however, demonstrated that these *remember* false alarms are often made faster and with higher confidence than *know* hits, and are correlated with correct *remember* judgments, both of which pose problems for theories assuming a metacognitively accessible recollection threshold.

Parks & Yonelinas (2007) acknowledge these limitations, and concede that *remember-know* reports do not provide process-pure estimates of recollection and familiarity. They also point to differences in instructions given in several studies (e.g., Stretch & Wixted, 1998; Rotello, MacMillan & Reeder, 2004). Some fail to instruct participants to respond with a *remember* judgment only if they can report the specific detail recollected about the test item. This may be a key factor in the presence of high levels of *remember* false alarms in some studies. Indeed, Rotello, MacMillan, Reeder, & Wong (2005) demonstrated that such differences in instruction could lead to higher levels of *remember* false alarms, potentially due to participants responding with *remember* judgments when experiencing high levels of familiarity. Parks & Yonelinas (2007) also

point out that *remember* false alarm rates are typically quite low, often falling between one and three percent.

Perhaps the most common criticisms directed toward the *remember-know* procedure involve whether an individual can metacognitively or consciously separate the two processes and make decisions based primarily upon one or the other. Wixted (2007) suggests that rather than assessing recollection, *remember* responses are in fact testing memory strength and confidence. If memories do fundamentally differ from *know* judgments, Wixted argues that this difference is one of conscious phenomenal experience of prior occurrence, or auto-noetic awareness, as Tulving (1985) originally proposed. This conscious experience may be only one aspect of the recollection process. Such an interpretation assumes that the type of subjective experience may be different for items that are *remembered*, but that the type of memory retrieved is not qualitatively different from that given a *know* response. This explanation may also account for an increasing amount of neuroimaging data that suggests a dissociation between brain regions involved in *remember* and *know* responses (e.g., Yonelinas, Otten, Shaw & Rugg, 2005; Eichenbaum, Yonelinas & Ranganath, 2007; Ranganath, Yonelinas, Cohen, Dyb, Tomb, D'Esposito, 2003).

However, Parks & Yonelinas (2007) point out that this explanation does not fully account for performance on other tasks, such as relational recognition tasks, that involve the same brain areas, but may not involve the same subjective phenomenal experience. The involvement of different brain areas in recollection and familiarity could potentially be explained by a signal detection interpretation conceding the presence of two

underlying memory processes. The dual-process UVSDT (Wixted, 2007) assumes that the brain areas differentially impact recollection or familiarity but effectively work in concert to produce the continuous memory strength signal described by signal detection theory. It is not clear, however, exactly how such an interpretation might be applied to existing imaging studies, which have typically been interpreted according to a dual-process perspective that assumes the two processes operate independently.

It is doubtful that any of the estimation methods outlined in this paper provide process-pure estimates of recollection and familiarity. Even dual-process theory's most vocal proponents (e.g., Parks & Yonelinas, 2007) distance themselves from such assertions. However, the convergence of these methods, as well as the support of the results of recent neuroimaging research suggest that these procedures are accessing the same two underlying processes. Additional manipulations and analyses are necessary to further elucidate the differences between these processes and their relative contributions to recognition memory. In particular, there is much to be learned about the encoding factors that lead to later differences in recollection or familiarity.

#### *The testing effect in recognition memory*

As mentioned previously, some convergence of process estimation results has been demonstrated using encoding manipulations. For example, increasing study duration tends to increase both recollection and familiarity similarly. However, having to generate an item at time of test leads to greater increases in recollection than familiarity. A relevant related question asks whether taking intervening tests between presentation of study materials and final test affects recollection and familiarity differentially. Research

on testing in memory has shown that taking a test over study material leads to greater retention than an additional study interval of approximately the same length, a phenomenon known as the testing effect. An abundance of literature has been compiled on this phenomenon and the effects of intervening tests on final recognition and recall (see Roediger & Karpicke, 2006 for review). Some evidence exists for differential effects of testing on recollection and familiarity, but an understanding of this evidence requires some knowledge of the origins of the testing effect, particularly involving recognition.

The basic idea behind the testing effect is far from novel, as knowledge of the beneficial effects of testing on memory for study materials predates William James, who postulated that “it pays better to wait and recollect by an effort from within, than to look at the book again” (p.646). Abbot (1909) was among the first researchers to scientifically study this effect. He replicated earlier studies by Witasek (1907; from Abbot 1909<sup>2</sup>) and Katzaroff (1908; from Abbot 1909<sup>3</sup>) that demonstrated superiority of study trials consisting of repeating non-sense words from memory over those consisting of simply re-reading the words. Gates (1917) and Spitzer (1939) further demonstrated the applicability of the effect in education, carrying out large scale studies on school children using stimuli ranging from articles to nonsense syllables.

Although effects of intervening recognition and recall tests on final recall ability have been consistently demonstrated (for review see Richardson, 1985; Roediger & Karpicke, 2006), consistent effects of intervening tests on final recognition tests have been more elusive. Hogan and Kitsch (1971) were among the first to directly compare



different types of intervening tasks and final recognition performance. They analyzed the effects of intervening two-alternative forced-choice recognition tests, free recall tests, or additional study list presentations on final tests consisting of either free recall or two-alternative forced-choice recognition. They found that intervening recognition and recall tests both resulted in better performance on final recall tests, but that study trials led to better performance on final recognition tests, regardless of the type of intervening test. Darley and Murdock (1971) also failed to find an effect of intervening recall tests when comparing performance on a final three alternative forced-choice recognition task and a final recall task. Participants who were tested with an intervening recall test did not differ in performance from those who were not tested when using a final recognition test.

Others have reported positive effects of testing on final recognition tests.

Hanawalt and Tarr (1961) examined testing effects using recognition final tests over incidentally encoded words (adjectives placed at the end of true-false questions). They found that intervening recall tests over these adjectives led to better recognition performance when tested 48 hours later, though non-testing conditions were not equated for exposure to study items. Cooper and Monk (1976) did equate for such exposure, and also reported a significant positive effect of recall testing on final recognition tests.

However, this effect was limited to a recall testing condition in which recall tests were alternated repeatedly with study trials, as final test performance in a condition featuring consecutive intervening recall tests following study trials did not differ from study-only conditions. Cooper and Monk also noted a shift to a more conservative response bias in

both intervening test conditions, though this shift was greater in the alternating test-study condition.

Minor changes in the methodologies of some earlier mentioned studies have also led to testing effects using final recognition tests. Lockhart's (1975) study was nearly identical to Darley and Murdock (1971), except faster presentation rates were used. Darley and Murdock presented words at a rate of five seconds per word, and Lockhart presented words at rates of 5s, 1500ms, or 750ms. Results suggested that effects of recall in Darley and Murdock's study were attenuated due to long presentation time, as testing effects were found in final recognition tests in both the 750ms and 1500ms conditions. However, this effect occurred only for items in the last few serial positions. Lockhart postulated that with presentation times as long as five seconds, recallable items will typically be recognized regardless of whether or not such a recall test is administered. Wenger, Thompson, and Bartling (1980) replicated Hogan and Kitsch's (1971) study and found that when exposure time was equated between intervening recall and no test conditions a significant positive effect of testing was obtained. However, this effect was somewhat inconsistent at shorter (10 minute) retention intervals.

Mandler and Rabinowitz (1981) also found prior testing to increase both final recall performance and recognition hit rates, though the increase in hit rates was overshadowed by a greater increase in false alarms. However, their list items and distractors consisted of items from similar semantic categories, and, as a result, the additional intrusions and false-alarms could potentially be an artifact of this methodology. In addition, many of the same distractor items were used for intervening

tests and final recognition tests, increasing the likelihood that these items would be later recognized.

Although the reasons behind early failure to find an advantage for intervening tests using recognition final tests are not well known, significant effects of intervening recall and recognition tests have been found using both recognition and recall final tests. Several explanations exist for the finding that taking a test over material typically improves memory to a greater extent than additional exposure. One such explanation is the transfer appropriate processing view (Morris, Bransford, & Franks, 1977). This view states that the processes involved in retrieval of information during an intervening test are similar to those used during a final test over the material. This view is supported by numerous studies demonstrating *encoding specificity*, which states that similarity of context at study and test lead to improved memory performance (e.g., Tulving & Thompson, 1971; Godden & Baddeley, 1975). Similarly, a match between the encoding operations performed during the intervening test and retrieval processes at final test leads to greater performance when compared with conditions in which these encoding operations involved additional study without such testing. Studies manipulating the types of cues given at study and test in cued recall tests have shown advantages for testing only when the cues at study and test matched (McDaniel, Kowitz, & Dunay, 1989; McDaniel & Masson, 1985). This transfer effect may also impact recall and recognition intervening and final tests. Duchastel & Nungester (1982) gave high school students intervening multiple choice or short answer tests over studied material. Although final test

performance was greater than controls regardless of intervening or final test format, final test performance was higher when intervening and final test formats were the same.

A second view, often referred to as the elaborative processing view (e.g., Carpenter & DeLosh, 2006), postulates that rather than invoking processes similar to those used on a final test, intervening tests result in more elaborative processing than study trials. Glover (1989) referred to such a view as the retrieval hypothesis, and demonstrated that the effect of intervening tests on final test performance was influenced by the level of elaboration required in the intervening test (free recall vs. cued recall vs. recognition). A greater amount of elaboration required during intervening test resulted in better memory performance on the final test. Glover (1989) also used free recall, cued recall, and recognition final tests, in a direct comparison of the elaborative and transfer appropriate processing views. He found that regardless of the type of final test, free recall intervening tests led to better memory performance. This is consistent with the elaborative view, but not with the transfer appropriate processing view, which predicts better performance when intervening and final test formats are similar. Carpenter and DeLosh (2006) recently replicated Glover's (1989) study using word lists as study materials and a shorter retention interval. Carpenter and DeLosh failed to find a significant advantage for cued recall or free recall intervening tests on final recognition performance. They also failed to find a significant advantage for any type of intervening test over a control study-only condition in several final test conditions. Kang, McDermott, and Roediger (2007), performed a similar study, in which participants studied brief journal articles and received multiple-choice or short answer intervening

tests or a list of pertinent statements to read. When no feedback was given, performance on later short answer and multiple choice tests was greater in intervening multiple-choice test conditions. Such studies provide only mixed support for an elaborative processing view, particularly in light of the consistent failure to find an advantage of ‘more elaborative’ intervening recall tests when using recognition final test formats. Exceptions have generally involved situations in which feedback was given following intervening tests, which seems to benefit recall tests more than recognition (e.g., McDaniel, Anderson, Derbish, & Morrisette, 2007; Kang, McDermott, & Roediger, 2007).

*The testing effect and the dual process approach*

Mandler and Rabinowitz (1981) were the first to refer to the dual process approach to recognition in attempting to explain their results. They postulated that the additional tests incremented the familiarity values of the studied items, but that the increase in familiarity value of lures also increased. This assertion is supported by the reduction in reaction times for studied items with each test, which is assumed to be caused by increased reliance on a faster familiarity process, resulting from these items exceeding the upper familiarity criterion and therefore foregoing a search process (see Figure 1). Re-testing of lures also led to increases in reaction times due to increased familiarity values. A possible explanation for this increase in reaction times is that retesting of these items increased their familiarity values enough to fall between decision criteria, therefore necessitating a search process.

Similarly, Chan & McDermott (2007) suggested that the failure of some studies to find significant effects of intervening tests on final recognition tests could be due to

intervening tests increasing recollection without impacting familiarity. Only two studies have attempted to assess the differential effects of testing on recollection and familiarity using process estimation methods as final tests. The first was conducted by Jones and Roediger (1995). They presented several lists of eight words each, with intervening recall tests over half of these words. A *remember-know* test was then administered after a short delay. Results indicated a greater hit rate for tested items, and that the source of this increase in hit rate was an increase in *remember* judgments. However, like Lockhart (1975), they found that this increase in hit rate, and corresponding increase in *remember* judgments, was restricted to the last few items presented on each list. This pronounced recency effect results from the immediate recall test following the list presentation, and was likely further facilitated by Jones and Roediger's instructions to first recall the words from the end of the list. Although this may have achieved its intended effect of "preventing subjects from adopting different recall strategies across lists" (p. 69) it may have also affected the contribution of recollection to these items.

Chan and McDermott (2007) conducted a similar study. In this study, participants studied four word lists and were given intervening free recall tests after two of these lists. The other two lists were followed by distractor tasks. In experiments 1a and 1b, a final source memory or exclusion task was used to assess the effects of intervening free recall tests on recollection and familiarity. Experiment 2 examined the same effect with a final *remember-know* task. Results indicated that taking an intervening recall test showed inconsistent effects on the hit and false alarm rates, as significant effects were found using an exclusion task but not *remember-know* or source memory tasks. However,

testing did increase correct source judgments and the probability of recollection-based responding in both the exclusion tasks and *remember-know* tasks. Although Chan and McDermott also examined serial position effects, unlike Jones and Roediger (1995) they reported an increase in recollection across serial positions, not merely the final positions. They posit the question of whether this effect is unique to intervening recall tests or could potentially be found using other types of intervening tests, as little is known about the effect of type of intervening task on recognition memory.

*An examination of the effects of type of intervening test on recognition final tests*

The current research is an attempt to determine the effect of type of intervening test on final recognition, as well as recollection and familiarity components. This research is not a direct attempt to test models of recognition, though results obtained may contribute to the literature explaining how different manipulations at encoding affect the two processes underlying recognition memory. Due to the aforementioned concerns about process estimation methods, and the lack of a single process-pure estimation method, this study used two commonly utilized estimation methods; a source memory task and a *remember-know* task. If differential effects of testing were found, this would provide an additional test of the convergence between methods, as well as ensure that any such effects are detected. Because a within-subjects design with several testing conditions was used, the potential effects of presentation order was also assessed.

Differences in intervening test performance were also analyzed. Participants were expected to correctly identify a higher proportion of study items using recognition intervening tests. Smith and Barker (under review) have shown advantages to using yes-

no tests to assess classroom based knowledge, including strong relationships to other indices of classroom performance. The effects of the additional alternatives used in multiple choice tests commonly found in classroom settings are not well established (but see Butler, Marsh, Goode, & Roediger, 2006). Verbal reports of participant preferences for the yes-no test (Cameron, 2002; Kojic-Sabo & Lightbown, 1999) as well as others suggesting the yes-no tests were more difficult (Yonelinas, Hockley, & Murdock, 1992) have been cited, but no clear agreement exists about which is the more sensitive measure. Some studies have demonstrated performance equivalence between yes-no and multiple choice performance (e.g., Green & Moses, 1966), others have suggested a performance advantage for multiple choice tests (e.g., MacMillan & Creelman, 1991). In several studies that suggested an advantage for multiple choice testing pictures were used as stimuli (e.g., Kroll, Yonelinas, Dobbins & Frederick, 2002; Deffenbacher, Leu & Brown, 1981). Other studies have found equivalent results between multiple choice and yes-no tests using as few as one foil per test item (i.e., a two-alternative forced-choice task; Yonelinas, Hockley, & Murdock, 1992). To more closely align with the type of multiple choice test used in the classroom, four alternatives were used in each multiple choice question in the present studies, three of which were foils. Yonelinas, Hockley, and Murdock (1992) report initially equivalent results using multiple choice and yes-no tests over study words, but also found that only subjects who were repeatedly tested using the yes-no format improved in performance across trials. The present studies should not only test the contribution of these test types to later performance, but also allow for a direct test of performance differences between these methodologies.



Of additional concern when using intervening and final recognition tests is the use of foils. Using the same foils for both intervening recognition test and final recognition test often results in a greater increase in the false alarm rate than hit rate (Hogan & Kitsch, 1971; Mandler & Rabinowitz, 1981; Richardson, 1985) presumably due to recognizing the foils from the earlier intervening test. However, using different foils at intervening test and final test has been shown to increase hit rate (Hogan & Kitsch, 1971; Richardson, 1985), which may result from memory of these items at test. Since memory for items at test is being directly compared with memory for re-studied items, any such advantages are seen as part of the beneficial effects of testing in the present experiments. In these experiments study items were re-presented in the no-testing condition. As a result, a direct comparison may be made between any improved discriminability due to re-presentation of studied items in an intervening test and the improved discriminability due to a mere re-presentation during the additional study trial. It must then be assumed that any performance difference between the two conditions at final test is due to testing, and not merely additional exposure.

Several questions involving performance at final test were addressed in this research. The first involves the effects of testing, and type of test, on overall performance at final test. Although several researchers have also shown that the type of intervening test used (i.e., recall, recognition, multiple choice) may differentially affect memory for studied materials at final test (Hogan & Kitsch, 1971; Duchastel, 1981; Glover, 1989; Kang, McDermott & Roediger, 2007; Butler & Roediger, 2007; Carpenter & Delosh, 2006), the findings regarding how these intervening test types impact final recognition

performance were quite inconsistent. In addition, none of these studies have examined these effects on recollection and familiarity. The present research is an attempt to better understand the effects of intervening test types, compared to re-presentation of the material without testing, on final recognition as well as its components. If some or all intervening tests lead to greater performance on final recognition tests than no-test conditions, this will provide additional support for the testing effect in recognition.

If intervening recall tests lead to greater performance than multiple-choice or yes-no recognition this will support the elaborative processing view, though the alternative would support a transfer appropriate processing view. Whether encoding processes during intervening recall, multiple-choice, and yes-no tests differ in their contribution to later recollection and familiarity assessments, and whether any of these conditions individually differed from a no-test condition in which words were presented for a second time were also assessed. Neither of the previous studies looking at process-estimation methods to assess recollection and familiarity (Jones & McDermott, 1995; Chan & Roediger, 2007) have attempted to equate exposure time in the no-test condition. At least one recent study (Carpenter & DeLosh, 2006) suggests that equating exposure for words in a single session may lead to greater overall recognition performance in non-tested conditions than some testing conditions. However, the effects on recollection and familiarity are not known. Finding that type of intervening test differentially impacts recollection and familiarity would be the first such demonstration in the literature. Final test performance in one or more intervening test conditions reflecting a greater proportion of *remember* or correct source judgments would support the assertion that intervening

tests differentially affect recollection and familiarity. If, however, no such differences exist, this may provide support for the more parsimonious notion that the advantage seen in earlier studies was likely due to additional exposure to the material, and is not specific to taking an intervening test. Finally, it may be that only items that are correctly identified at intervening test are more likely to be recollected at final test. This would indicate that intervening test performance, and not simply the act of taking such a test, differentially influences recollection and familiarity.

Additionally, differences in responding criterion at final test across testing conditions were determined. Some research suggests that responding becomes more conservative with testing over the material (Cooper & Monk, 1976) and other research suggests that additional study time also leads to more conservative responding strategies (Ruiz, Soler & Dasi, 2004). More conservative responding strategies have been found to correlate strongly with performance (e.g., Smith, 2006), and Feenan & Snodgrass (1990) cautioned against simply treating bias as a nuisance variable, stating that it is important to understand this part of the memory process and how it is manifested in recognition performance. Finding that participants adopted more conservative responding strategies following testing conditions than following additional study time, or that differences in responding criterion depending on type of intervening test, would provide insight into this relatively unexplored component of recognition tests. Failure to find differences may suggest that study and test trials lead to comparable changes in responding criterion.

Finally, response time data for final source and *remember-know* judgments were obtained. As mentioned previously, familiarity-based responses are assumed to be the

result of a faster process than recollection based responses, though *remember-know* studies have found the reverse effect. Response times to items correctly *remembered*, or given correct source judgments, may be different than those later given *know* responses, or incorrect or no-source judgments. Longer response times for *remembered* items would support numerous previous findings that suggest faster *remember* judgments than *know* judgments in the *remember-know* paradigm, though this contradicts the basic notion that familiarity is a faster process than recollection. Finding convergent results using the source memory tests (i.e., faster source memory judgments than no source judgments) would support the notion that these response time contradictions likely result from participants first searching for recollective detail when this information is requested as part of the testing methodology.

## II. EXPERIMENT 1

### *Method*

#### *Participants*

Undergraduate students at Auburn University ( $n = 64$ ) participated in return for course credit and \$10.00 cash. Participants were tested in groups of 10 or fewer. The data from three participants were not included in these analyses due to response times less than 500 milliseconds for multiple test items<sup>4</sup>. Twenty-nine participants remained for analysis in the *remember-know* final test condition, thirty-two in the source memory final test condition.

#### *Materials*

Four-hundred and eighty unrelated words (see Appendix B) were selected from the MRC Psycholinguistic Database based on the following criteria; number of syllables (1-3), number of letters (3-7), Kucera-Francis (1967) frequency (20 to 100 occurrences per million), and concreteness score (between 200-600; see Appendix C for all word list means). Words were then randomly assigned to one of twenty-four lists consisting of 20 words each for counterbalancing purposes. Word lists were equivalent in syllables ( $M = 1.51$ ,  $SD = 0.11$ ), letters ( $M = 5.17$ ,  $SD = 0.24$ ), frequency ( $M = 47.26$ ,  $SD = 5.03$ ), concreteness ( $M = 456.32$ ,  $SD = 19.11$ ), and imagery ( $M = 487.35$ ,  $SD = 14.76$ ). Eight lists were used as study lists for each of the four intervening test conditions. The words

on the remaining sixteen lists were used as foils in the multiple-choice, yes-no, and process-estimation recognition tests. Stimuli were presented on 10.5 x 13.25 inch flatscreen monitors using Intel ® Core 2™ 2.4 GHz desktop computers equipped with Superlab 4.0® stimulus presentation software. Study words were presented in the center of the screen (see Appendices D and E for sample instructions and test screens, respectively). Participants responded by using the computer keypad.

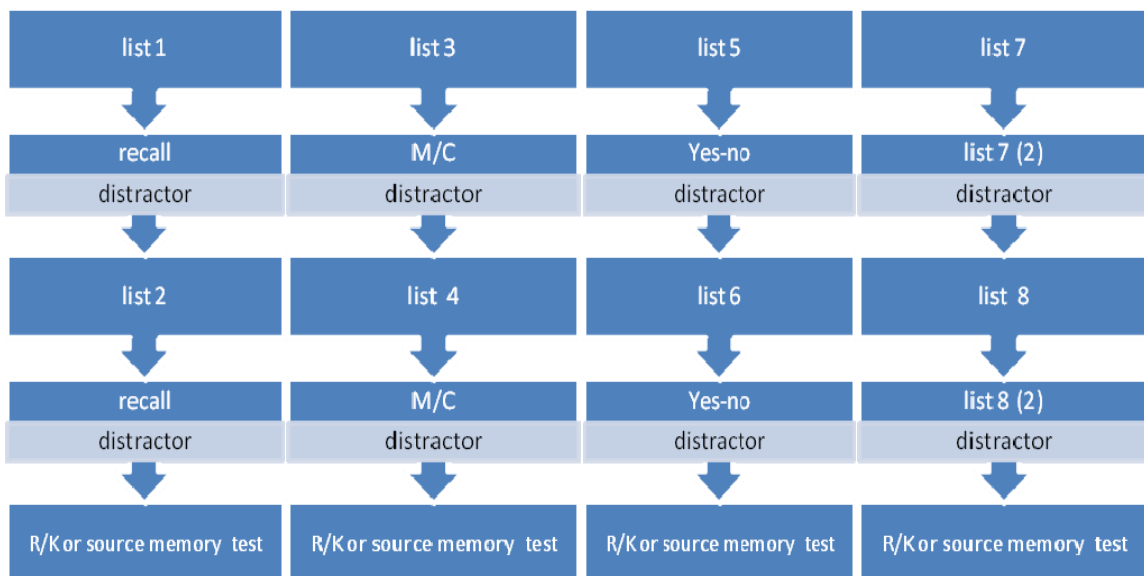


Figure 3. Design for Experiment 1.

### Design

Experiment 1 contained a within-subjects manipulation (type of intervening test: recall, multiple-choice, yes-no, and no-test) and a between subjects manipulation (source or *remember-know* final process estimation tests). Participants were exposed to four intervening test conditions, each consisting of two randomized study lists, an intervening

test over each list, and a final process-estimation test (see Figure 3). The order of intervening test condition (multiple-choice, recall, yes-no, or no-test), and the order of lists presented (8 lists) were counterbalanced using balanced Latin square designs (Williams, 1949). Final process-estimation procedure (source memory or remember-know) was randomly assigned to experimental session.

### *Procedure*

Participants self-scheduled to an available 1.5 hour session. Upon arrival they were instructed to sit at a computer. Experimental conditions, such as order of intervening test conditions and word lists, were randomly assigned to computers before their arrival. Five to nine participants were tested per session. Participants were first informed of the structure of the experiment and given instructions on how to respond in the final process estimation procedures (see Appendix D for sample instructions). These instructions served two purposes beyond informing them that all tasks would be completed on their computers. First, instructions in the *remember-know* conditions served to clarify the types of memory experiences that should lead to *remember* or *know* responses. Second, they ensured that participants in the source memory task approached all four conditions with the knowledge that they would be tested on the source of the presented words. This was deemed preferable to participants discovering this at the end of block one, which could potentially change the way participants attended to or encoded items for the remaining three blocks.

During each study list, 20 words were individually presented for 2s each, separated by a blank screen for 500ms. Each study list lasted 50s. The study lists were

followed by an intervening test or no-test condition, which lasted approximately 90s. Multiple-choice questions consisted of one studied item and three new items, each designated a number (1-4; see Appendix E for screenshots of sample test items). Participants were asked to select the number of the item that was studied on the previous list. Yes-no tests consisted of sequential randomized presentation of the twenty studied items and twenty new items. Participants were asked to press the '1' key if the item was on the previous list and the '2' key if it was not. Free recall instructions were simply to type as many of the items from the previous list as the participant could remember. Following the first intervening test, participants were given a 30s Brown-Peterson distractor task, which involved counting backwards by threes from a three digit number. Participants typed these responses on their computer keyboard to ensure their cooperation. This was followed by the presentation of the next study list of twenty words, a test of the same format over the second study list, and another 30s distractor task of the same type. After intervening tests over both lists were completed, participants were presented with an additional distractor task for 180s. This distractor task consisted of a recognition test over the numbers that the participant had previously generated during earlier distractor tasks as well as the presentation of an equal number of basic math equations that asked the participant to decide whether the answer provided was correct. Responses to the distractor tasks were not recorded.

To control for differences in the amount of time required to complete each type of test, the interval between test items was held constant during the multiple-choice (750ms) and yes-no (1310ms) tests. These times were based on pilot data that indicated that with



an interstimulus interval of 0s the mean time to complete each multiple-choice and yes-no intervening tests were 75.5s and 37.6s, respectively. During these intervening tests, each response to an intervening test item was followed by a blank screen for the designated interval. Participants were given 90s to type their responses in the recall test condition. In the no-test condition participants were presented with a study list followed by a second presentation of the same study items in random order at the same 2s rate. This was followed by an additional 70s Brown-Peterson distractor task. The same procedure was then followed for the second list, followed by the final 180s number recognition distractor task outlined above.

At the completion of each condition participants were given a process estimation test. Some participants received a *remember-know* final test (N = 31), others received a source memory final test (N = 33). Both final tests consisted of sequential presentation of study words and non-presented lures. Final test type was randomly assigned to experimental session. Participants given a *remember-know* task as the process-estimation procedure were asked to (a) “press the “3” key if you can *remember* specific details associated with the word’s presentation during the study episode, and could report these details,” (b) “press the “4” key if you can not specifically *remember* details of the item’s presentation but the word is familiar, and you know it was presented,” or (c) “press the “5” key if the item was not a studied word.” These instructions have been suggested by Parks & Yonelinas (2007) based upon previous research (Rotello, McMillan, Reeder & Wong, 2005) that suggests that slight differences in instructions can influence false *remember* judgments. Parks and Yonelinas suggest that failure to instruct participants to

respond *remember* only if they could report what was recollected about the item at study if asked could lead to participants also responding *remember* when the item has a high familiarity value.

Participants given the source memory task were asked to make one of four responses. Participants were instructed to (a) “press the “1” key if the item was on the first study list,” (b) “press the “2” key if the item was on the second study list,” (c) “press the “3” key if the item was studied but you can not *remember* which list it was on,” or (d) “press the “4” key if the item was not part of a study list.”

Response times were collected, though participants were not given instructions to respond as quickly as possible (i.e., non-speeded conditions). This was due to this study’s emphasis on the testing effect and type of memory experience, and a desire not to artificially compress such experiences. This was particularly relevant in light of pilot data that suggested that such instructions could lead to a propensity to respond too quickly to make appropriate decisions about these experiences (i.e., response times less than 500ms). Instructions directed participants to “take the time needed to respond with your best answer, and no more” (see Appendix D). A one-step procedure was used, in which participants were asked to make only one decision (i.e., *remember, know, or no*), rather than asking them to make a yes-no decision followed by a *remember-know* or source decision. This should, however, serve to more closely approximate typical testing situations, and allow for the best chance of response time convergence between methods. In both process estimation tests participants are likely to wait until recollective or source information is available before responding.

## *Results & Discussion*

### *Intervening test performance*

All analyses involve submitting intervening test conditions to repeated one-way ANOVAs unless otherwise stated. The alpha level for all analyses was set at .05. The proportion of items successfully retrieved or recognized was computed for all intervening tests. Because the intervening test conditions were identical for all participants, intervening test scores for both conditions were submitted to a single one-way repeated measures ANOVA. The results of the ANOVA showed a significant main effect of test type,  $F(2, 118) = 480.07, p < .01$ , and post hoc comparisons using Bonferroni's correction showed that performance on yes-no intervening tests ( $M = .86, SD = .08$ ) was significantly better than multiple choice tests ( $M = .78, SD = .17$ ) and recall tests ( $M = .30, SD = .13$ ) performance. The difference between the latter two test types was also significant. An interaction between conditions was also found,  $F(2, 118) = 5.06, p < .01$ , likely the result of higher performance on the intervening multiple-choice tests in participants in the final source memory test condition, ( $M = .84, SD = .14$ ) than those in the final *remember-know* test condition ( $M = .71, SD = .19$ ). The finding that performance in the two recognition tests was higher than performance in the recall task (see Figure 4) is not surprising. However, the reasons behind the increased multiple-choice performance in the source memory final test condition are unknown, as this is the only condition in these analyses in which multiple choice intervening test did not differ significantly from yes-no performance.

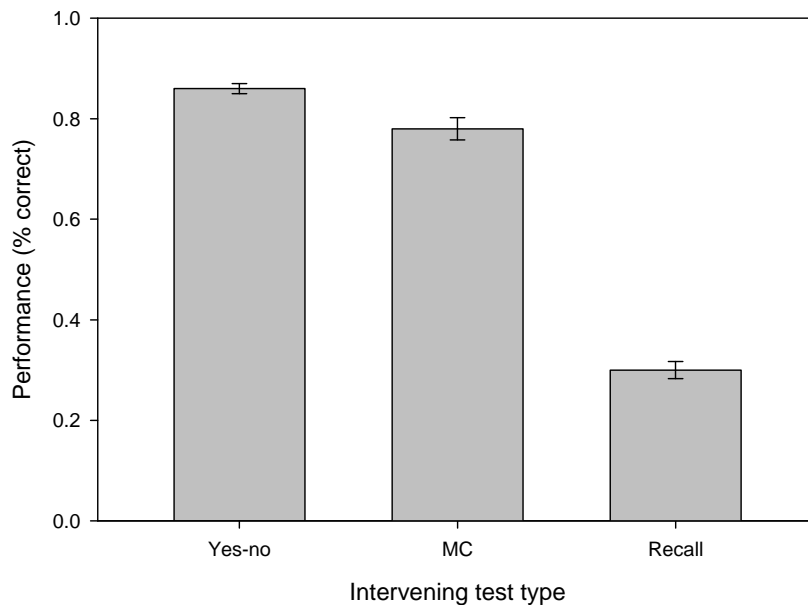


Figure 4. Mean Intervening Test Performance by Test Type.

#### *Final process-estimation test performance*

Final process-estimation test performance using *remember-know* and source memory tests was calculated using both hits minus false alarms and signal detection measure  $d'$ . Because these calculations yielded similar results, and because hits minus false alarms is a commonly used correction for guessing in recognition memory studies, hits minus false alarms were used in subsequent analyses. Any item correctly identified as having been previously presented, regardless of knowledge of the item's source or whether it was given a *remember* or *know* judgment, was scored as a hit (see Table 1 for total hits and false alarms for both conditions). When determining a hit rate, the variable of interest was whether the item was correctly identified as a studied item. Additional

analyses examined the exact nature of the correctly identified studied items. Any response to a non-studied item aside from correctly identifying it as such was scored as a false alarm. Because the final *remember-know* and source memory tests ask participants to make different decisions, all results obtained using these final estimation procedures were analyzed separately.

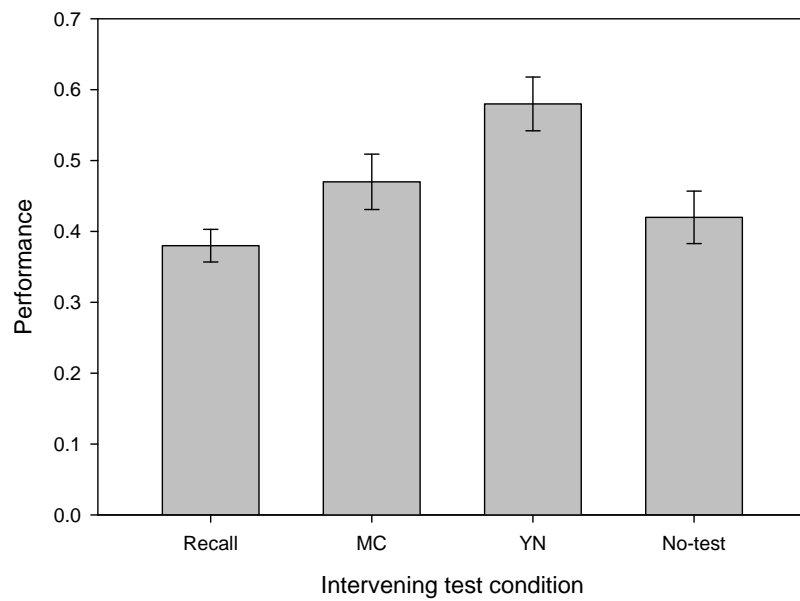
Table 1

*Final Performance Rates for Remember-Know and Source Memory Conditions.*

Variable	Condition			
	Recall Mean (SD)	MC Mean (SD)	YN Mean (SD)	No-test Mean (SD)
Remember-know condition				
Overall	.38 (.12)	.47 (.21)	.58 (.21)	.42 (.20)
False alarm rate	.43 (.11)	.38 (.19)	.31 (.20)	.41 (.18)
Hit rate	.81 (.11)	.84 (.09)	.90 (.09)	.84 (.13)
Source memory condition				
Overall	.49 (.21)	.64 (.24)	.61 (.23)	.58 (.25)
False alarm rate	.33 (.23)	.25 (.23)	.27 (.24)	.29 (.24)
Hit rate	.83 (.15)	.89 (.14)	.88 (.11)	.87 (.13)

A repeated measures ANOVA of overall final *remember-know* test performance demonstrated significant differences between intervening test conditions,  $F(3, 75) = 10.20, p < .01$ . Planned post-hoc comparisons using Bonferroni's correction showed that final recognition performance was significantly higher in the yes-no intervening test condition ( $M = .58, SD = .21$ ) than the recall intervening test condition ( $M = .38, SD = .12$ ), and the no-test condition ( $M = .42, SD = .20$ ). Performance in the multiple choice intervening test condition ( $M = .47, SD = .21$ ) was higher than in the no-test and recall conditions (see Figure 5), but these differences did not reach significance using

Bonferroni's correction. Performance differences reflected both an increase in hits and a decrease in false alarms for the yes-no testing condition (see Figure 6).



*Figure 5.* Comparison of Intervening Test Conditions on Final *Remember-Know* Test Performance.

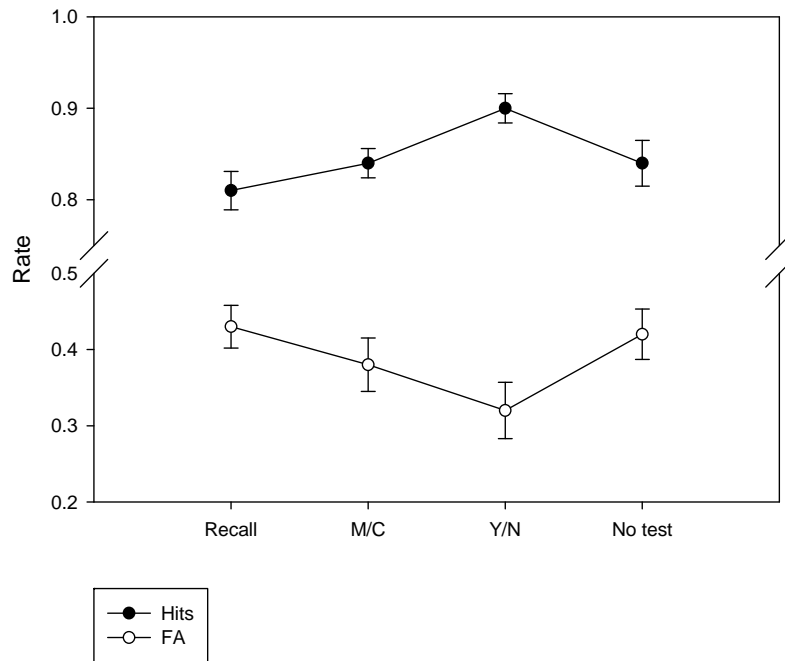
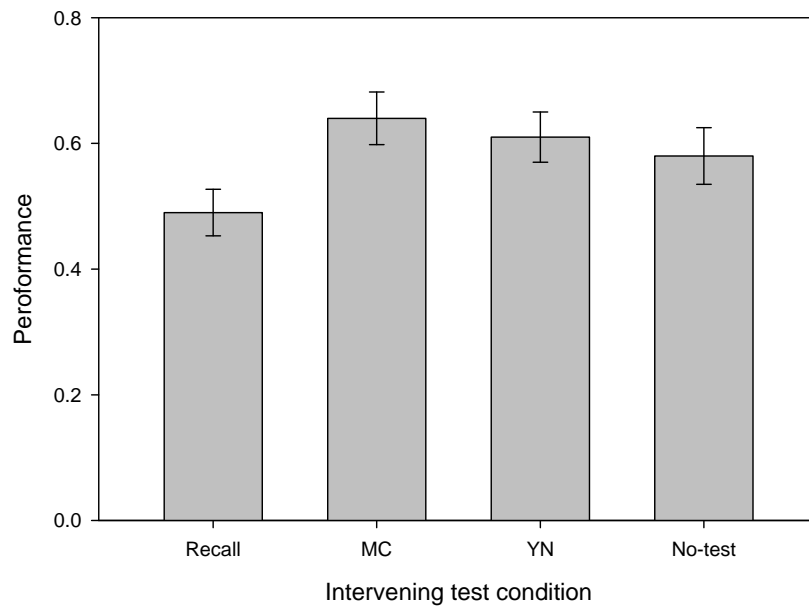


Figure 6. Comparison of the Contribution of Hits and FA to Performance by Intervening Test Condition in Final *Remember-Know* Test Condition.

The ANOVA for overall source memory test performance yielded a significant result,  $F(3, 84) = 8.25, p < .01$ . Post-hoc comparisons showed that final performance was significantly lower in the intervening recall test condition ( $M = .49, SD = .21$ ) than performance in the multiple choice condition ( $M = .64, SD = .24$ ) and the yes-no condition ( $M = .61, SD = .23$ ). As seen in the *remember-know* condition, the multiple choice and yes-no testing conditions slightly outperformed the no-test condition ( $M = .58, SD = .25$ ; see Figure 7), however, this difference did not reach significance. The results of the source memory condition are the only results in these analyses that did not

yield a significant testing effect. Possible explanations for this finding will be discussed shortly.



*Figure 7.* Comparison of Intervening Test Conditions on Final Test Performance in Source Memory Condition.

List order, which was counterbalanced using a Latin Square design, was included as a between-subjects variable in the preceding analyses to determine whether overall performance was influenced by position of the intervening test condition in the analysis. The interaction between the test type and order was non-significant in both the remember-know condition,  $F(9, 75) = .90, p = .53$ , and the source memory condition,  $F(9, 84) = 1.65, p = .114$ .



Table 2

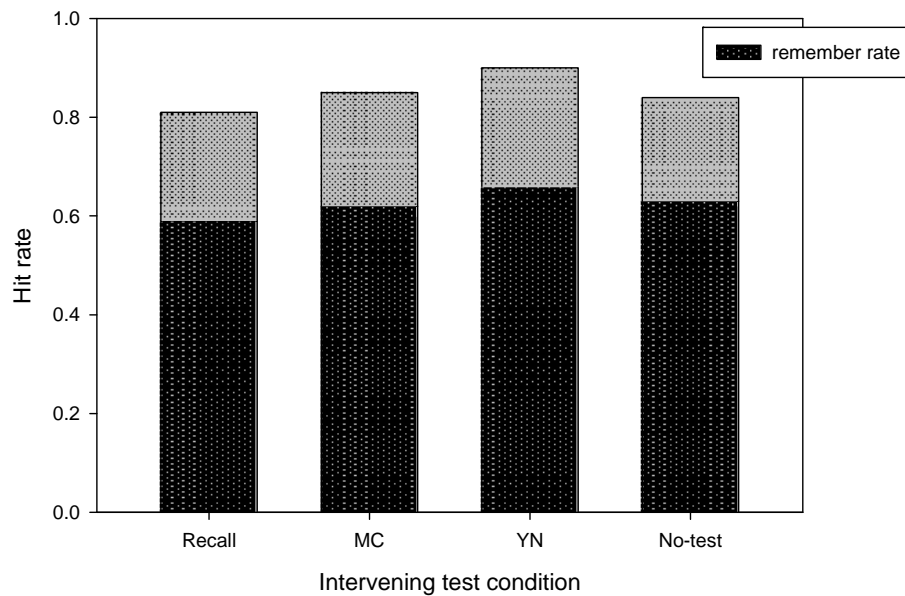
*Final Recollection Probabilities for Remember-Know and Source Conditions.*

Variable	Condition			
	Recall Probability (SD)	MC Probability (SD)	YN Probability (SD)	No-test Probability (SD)
Remember-know condition				
Raw remember	.59 (.12)	.62 (.19)	.66 (.18)	.63 (.22)
Conditional remember	.71 (.15)	.72 (.18)	.73 (.16)	.73 (.20)
Source memory condition				
Correct source	.42 (.17)	.37 (.21)	.46 (.21)	.44 (.24)
Incorrect Source	.21 (.13)	.23 (.16)	.23 (.14)	.18 (.13)
No source	.21 (.14)	.29 (.21)	.21 (.18)	.28 (.18)

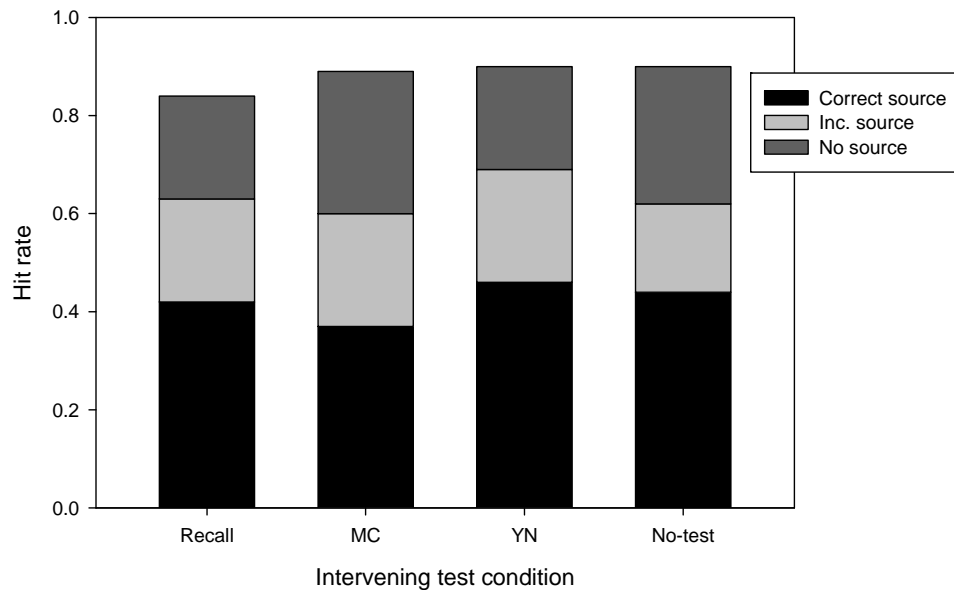
*Recollection and familiarity across conditions*

The *remember-know* and source memory final process estimation tests were also designed to determine the participant's type of memory experience. Because the final *remember-know* and source memory tests may not be measuring the same underlying constructs, as mentioned above, they were analyzed separately. To determine the potential effect of intervening test type on participants' *remember* judgments and memory for source, the proportion of *remember* judgments and correct source judgments were submitted to repeated measures ANOVAs. Neither the repeated measures ANOVA for the *remember-know* condition,  $F(3, 84) = 1.33, p = .27$ , nor the repeated measures ANOVA for the source memory condition,  $F(3, 93) = 1.58, p = .20$ , were significant. The raw and conditional probabilities of *remember* judgments and correct source judgments for each intervening test condition are listed on Table 2. It does not appear that taking an intervening test of any type leads to greater recollection than additional study trials when assessed with *remember-know* or source memory estimation procedures. This

was true regardless of whether raw remember probabilities or conditional probabilities of remembering (given the item was correctly identified at final test) were used. Although the proportion of *remember* judgments did not differ across intervening test conditions (see Figures 8 and 9 for the contribution of recollection-based responding to the hit total), the raw probabilities of *remember* judgments in all intervening test conditions (Ms .59 to .66) were higher than those in Chan and McDermott's (2007) no-test condition (M = .44). This suggests that their reported recollection advantage in recall intervening test conditions may be due to additional exposure during testing, rather than the encoding effects of the test, as exposure to items was not equated in their no-test condition.



*Figure 8.* The Contribution of *Remember* and *Know* Responses to the Overall Final Test Hit Rates by Intervening Test Condition.



*Figure 9.* The contribution of correct source (dark), incorrect source (light), and no source (medium) judgments to the overall final test hit rates in by intervening test condition.

*Recollection for items correctly identified at intervening test*

Whether a recollection advantage could be found for items that were correctly identified at intervening test, relative to items not correctly identified at intervening test and items in the no-test condition was also addressed. Although no advantage for recollection was found in any testing condition, this could be due to failure for items that were missed at intervening test to be enhanced by testing. Three separate repeated measures ANOVAs were conducted for remember-know final test conditions. Conditional probabilities, the probabilities of recollection at final test given performance at intervening test, were used in these analyses. The variable of interest in these analyses

was proportion of recollection-based responses for items that were correctly identified at intervening test.

The first such analysis consisted of conditional *remember* probabilities for items that were correctly recalled at intervening recall test, items that were missed at intervening recall test, and items that were in the no-test condition. This ANOVA was significant,  $F(2, 56) = 73.7, p < .01$ . Post hoc comparisons using Bonferroni's correction indicated that items that were correctly recalled at intervening test were more likely to be *remembered* ( $M = .91, SD = .09$ ) than items that were not tested ( $M = .63, SD = .22$ ) or items that were not recalled at intervening test ( $M = .47, SD = .22$ ). The difference between nontested items and items not recalled at intervening test was also significant. The second ANOVA consisted of items that were correctly identified at intervening multiple choice test, items that were not identified at multiple choice test, and nontested items. This ANOVA was also significant,  $F(2, 54) = 30.5, p < .01$ . Post hoc comparisons indicated that items that were correctly identified at multiple choice test were also more likely to be *remembered* ( $M = .73, SD = .17$ ) than items that were not correctly identified ( $M = .34, SD = .29$ ) or items that were not tested ( $M = .62, SD = .22$ ). The difference between the later two groups was also significant. An ANOVA for items correctly identified at intervening yes-no test, items missed at intervening yes-no test, and nontested item was also significant,  $F(2, 54) = 26.2, p < .01$ . The same pattern was observed in post hoc comparisons as in multiple choice intervening tests, as items not correctly identified at yes-no intervening test were less likely to be *remembered* ( $M = .32,$

SD = .29) than items that were correctly identified at intervening test (M = .71, SD = .18) or items that were not tested (M = .63, SD = .22).

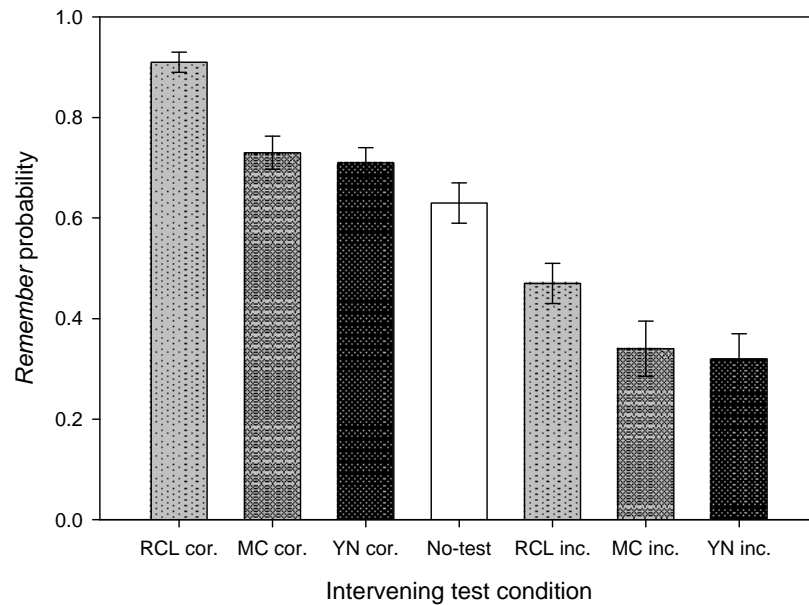
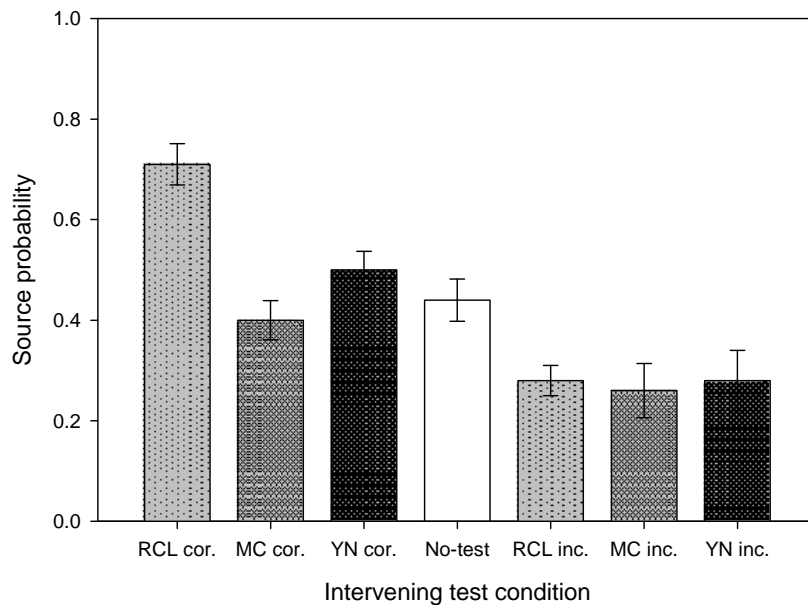


Figure 10. Mean Conditional Probabilities of remembering an Item at Final Test if the Item was Correctly Identified or Missed at Intervening Test or Not Tested.

Items that were correctly identified, missed, or not tested were also analyzed in source memory final test conditions. This was to determine if conditional probability of correct source judgment, given the item was correctly identified as a study item at final test, differed by intervening test status. An ANOVA for items that were correctly recalled or missed at intervening recall test and items that were not tested yielded significant results,  $F(2, 62) = 33.45, p < .01$ . Post hoc comparisons using Bonferroni's correction

showed that correct source judgments were more likely for items that were recalled at intervening test ( $M = .71$ ,  $SD = .23$ ) than items that were not recalled ( $M = .28$ ,  $SD = .23$ ) or nontested items ( $M = .44$ ,  $SD = .24$ ). The difference between the later two conditions was also significant. The ANOVA for items in the multiple choice condition that were correctly identified or missed and non-tested items was also significant,  $F(2, 52) = 4.99$ ,  $p = .01$ . Post hoc comparisons indicated that only the difference between items that were missed in intervening multiple choice tests ( $M = .24$ ,  $SD = .30$ ) and items that were not tested ( $M = .44$ ,  $SD = .24$ ) was significant. An ANOVA for items that were correctly identified or missed at intervening yes-no test and items that were not tested was also significant,  $F(2, 60) = 7.45$ ,  $p < .01$ . Post hoc comparisons indicated that items that source memory was more likely to be available for items that were correctly identified at intervening yes-no test ( $M = .50$ ,  $SD = .21$ ) than items that were missed at intervening test ( $M = .27$ ,  $SD = .35$ ). Neither differed significantly from nontested items ( $M = .44$ ,  $SD = .24$ ).



*Figure 11.* Mean Conditional Probabilities of Correct Source Judgment of Item at Final Test if the Item was Correctly Identified or Missed at Intervening Test or Not Tested.

Based on these analyses, it appears that correctly identifying an item at intervening recall test leads to a greater probability that it will be recollected at final test than additional exposure to the study items<sup>5</sup>. Correctly identifying an item at an intervening test of any type leads to a greater likelihood of recollection than was found for items that were not correctly identified, though this advantage is not consistently greater than additional exposure to the items. The recollection advantage for testing is in this way dependent on both the type of test and correctly identifying the item at test (see Figures 10 and 11). This is also true for the testing effect in overall performance, as items that are correctly identified at intervening test are better remembered at final test (see

Table 3 for a comparison of items that were correctly identified or missed at intervening test).

Table 3

*Final Test Performance Comparison of Items Correctly Identified or Missed at Intervening Test.*

<b>Final test</b>	<b>Intervening test</b>	
	<b>Correctly identified items</b> <i>Mean</i>	<b>Missed items</b> <i>Mean</i>
<u>Remember-know condition</u>		
N	33.17	46.83
Miss rate	.08	.25
Hit rate	.92	.75
<i>Remember</i>	.75	.41
<i>Know</i>	.17	.34
<u>Source memory condition</u>		
N	40.50	39.50
Miss rate	.06	.28
Hit rate	.94	.72
Correct source	.50	.25
Incorrect Source	.23	.21
No source	.21	.26

*Responding criterion*

Responding criterion was measured using the index C (Snodgrass & Corwin, 1988). Negative values of C represent liberal responding criterion, or propensity to respond that an item was previously presented. Positive values represent conservative responding, or a propensity to respond that an item is new. C was calculated for each participant's final process estimation tests, and submitted to repeated measures ANOVAs for remember-know and source memory final test conditions. ANOVA results did not reach significance in either *remember-know*,  $F(3, 84) = .60, p = .61$ , or source memory



conditions,  $F(3, 93) = .45, p = .72$ . These data do not support the hypothesis that presence, or type, of intervening test differentially affects criterion for responding. Although participants may change responding strategies across conditions or lists (e.g., Stretch & Wixted, 1998), it does not appear that these strategies change as a result of taking an intervening test (see Figure 12).

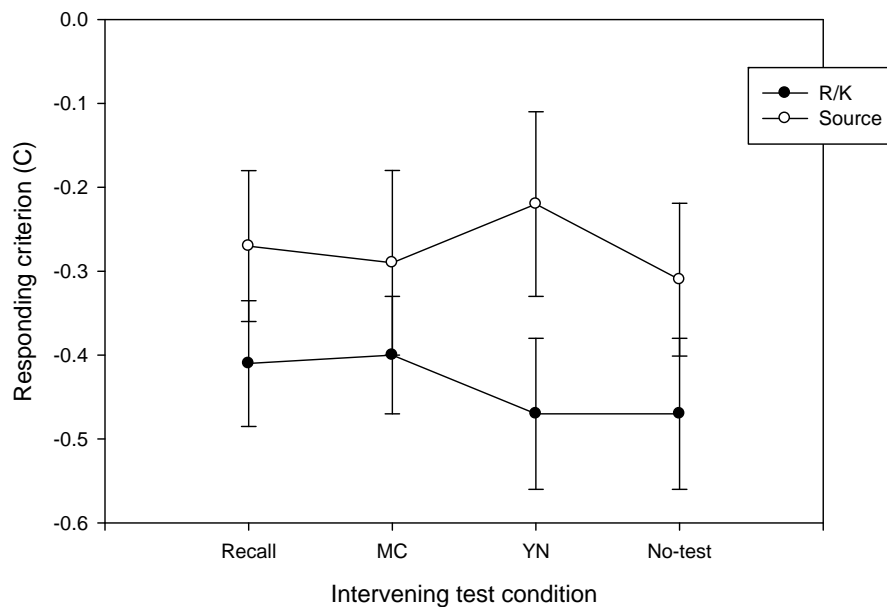


Figure 12. Responding Criterion (C) by Intervening Test Condition.

### Response times

Because response times were positively skewed, analyses were completed on raw response time as well as log transformed response times. Because the pattern of significant results was the same for both, raw response times are reported here. Response times to study items that were given *remember* and *know* judgments and foils that were incorrectly given *remember* judgments (*remember* false-alarms) were submitted to a

repeated measures ANOVA. The results of the ANOVA were significant,  $F(2, 56) = 31.50, p < .01$ . Post-hoc comparisons using Bonferroni's correction indicated response times for *remember* responses ( $M = 1517\text{ms}, SD = 587\text{ms}$ ) were faster than for *know* responses ( $M = 2536\text{ms}, SD = 1343\text{ms}$ ). As predicted by signal detection models, but not threshold models, *remember* false alarms ( $M = 1837\text{ms}, SD = 763\text{ms}$ ) were made significantly faster than *know* hits, but slower than correct *remember* responses (see Figure 13). This supports Wixted & Stretch's (2004) suggestion that *remember* and *know* judgments may be based on memory strength rather than a threshold-based recollection response.

An ANOVA for response times for correct source judgments, incorrect source judgments, and no-source judgments revealed a significant effect of source memory judgment on response time,  $F(2, 62) = 14.67, p < .01$ . Post-hoc comparisons using Bonferroni's correction revealed significantly slower response times for no-source judgments ( $M = 2785\text{ms}, SD = 1164\text{ms}$ ) than correct source judgments ( $M = 2228\text{ms}, SD = 723\text{ms}$ ) or incorrect source judgments ( $M = 2323\text{ms}, SD = 769\text{ms}$ ), with no significant difference between the later two judgment types (see Figure 13).

The finding that participants respond more quickly with correct source judgments than incorrect source judgments is likely an artifact of the use of a one-step procedure in this experiment. However, the finding that incorrect source judgments are made as quickly as source judgments and more quickly than correct no-source judgments, is more difficult to account for. Faster responding for incorrect source judgments than correct no-source judgments is, however, consistent with the *remember-know* findings. These results

suggest that, within a one-step process estimation procedure, responses based on incorrect recollective detail are made more quickly than responses without such recollective detail. This effect is not specific to the *remember-know* procedure.

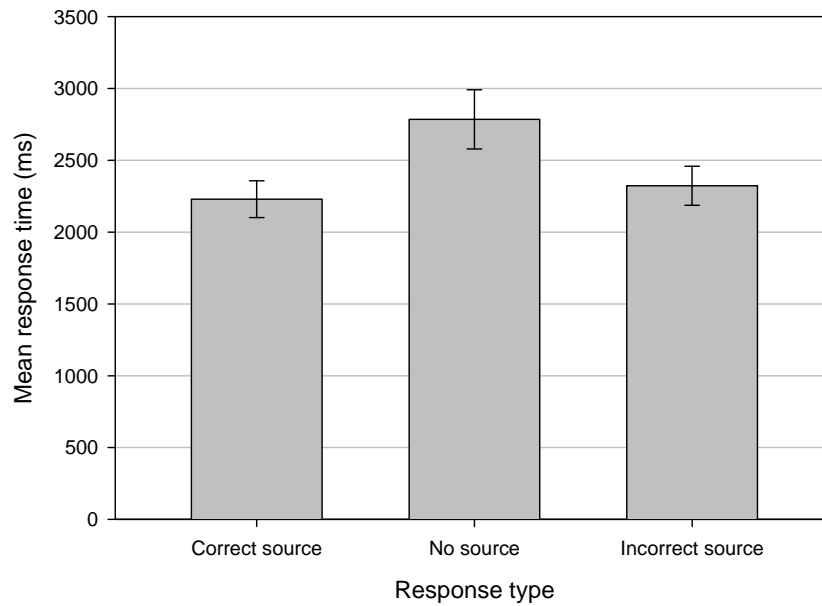
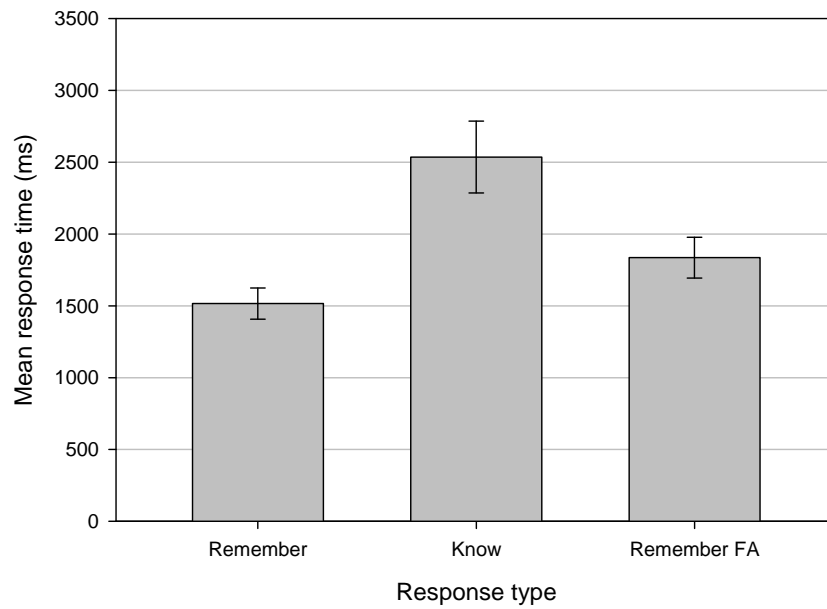


Figure 13. Response Time Differences For Correctly Recollected Items, Familiar Items, and Falsely Recollected Items in *Remember-Know* and Source Memory Conditions.

### III. EXPERIMENT 2

The high hit rates in Experiment 1 (see Table 1) suggest possible ceiling effects. This may have clouded the initial comparison of proportion of recollection-based responses between conditions at final test. Experiment 2 attempted to control these rates by including a distractor task immediately after presentation of each study list. The materials and procedure for the second experiment were identical to the first except for the addition of this distractor task. The additional distractor task used was a 30-second Brown-Peterson task, similar to Experiment 1. The distractor task was followed by the same intervening test that immediately followed the study list in Experiment 1. As in the first experiment, the final 180s distractor tasks consisted of a number recognition test and an equal number of basic math equations. Like the first experiment, both remember-know and source memory tests were used as final process estimation tests.

#### *Method*

##### *Participants*

Undergraduate students at Auburn University ( $n = 71$ ) participated in return for course credit and \$10.00 cash. Participants were tested in groups of 10 or fewer. Participants who completed Experiment 1 were precluded from participating in this second study. The data from two participants were not included in these analyses due to response times less than 500ms for multiple test items<sup>4</sup>. Thirty-four participants remained for analysis in remember-know conditions, thirty-five in source memory conditions.

## Design

The design used was identical to that used in Experiment 1 with the inclusion of a distractor task immediately following the study list (see Figure 14)

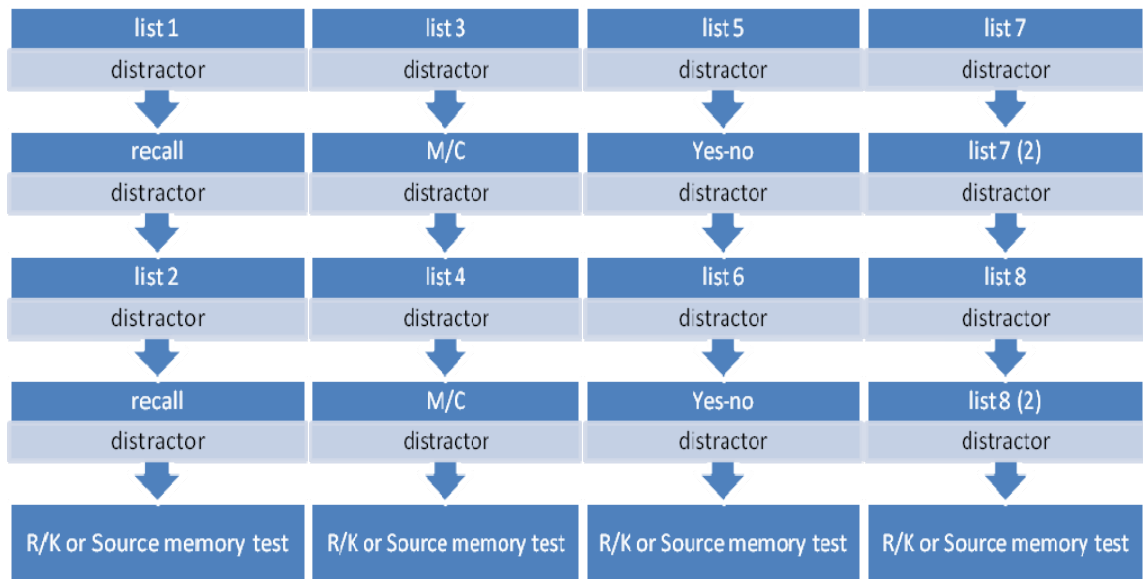


Figure 14. Design for Experiment 2.

## Results & Discussion

Inserting a distractor task immediately following study list decreased overall final process estimation hit rate across both conditions from .86 to .83. This reduction in hit rates was accompanied by an increase in false alarms and corresponding decrease in overall performance in source memory final test conditions. Curiously, in remember-know final test conditions in Experiment 2 the false alarm rate was lower than Experiment 1, resulting in a slight increase in performance. *Remember* judgments decreased in all conditions to levels comparable to Chan and McDermott's (2007) study,

though correct source rates also declined slightly. See Tables 1 and 4 for performance differences between experiments.

Table 4

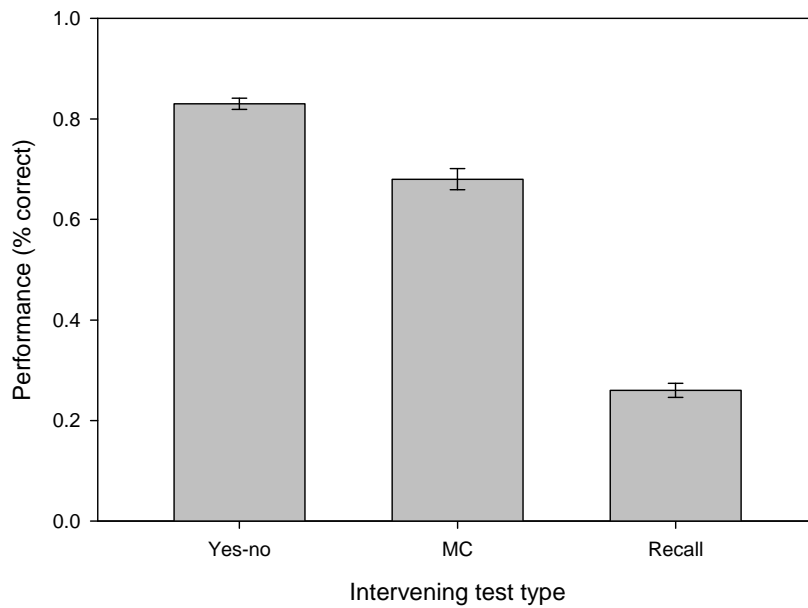
*Final Recognition Probabilities for Remember-Know and Source Conditions.*

Variable	Condition			
	Recall Mean (SD)	MC Mean (SD)	YN Mean (SD)	No-test Mean (SD)
Remember-know condition				
Overall	.43 (.18)	.56 (.19)	.61 (.19)	.49 (.19)
False alarm rate	.34 (.17)	.30 (.18)	.26 (.17)	.34 (.17)
Hit rate	.77 (.14)	.86 (.10)	.87 (.10)	.83 (.10)
Source memory condition				
Overall	.43 (.18)	.47 (.21)	.54 (.16)	.44 (.19)
False alarm rate	.39 (.21)	.38 (.20)	.32 (.18)	.39 (.17)
Hit rate	.82 (.11)	.85 (.11)	.86 (.12)	.83 (.11)

*Intervening test performance*

To test for differences in intervening test performance, intervening test scores were submitted to a repeated measures ANOVA. The results of the ANOVA were significant,  $F(2, 134) = 443.3, p < .01$ . Post-hoc comparisons using Bonferroni's correction showed higher performance on intervening yes-no ( $M = .83, SD = .09$ ) than multiple choice ( $M = .68, SD = .17$ ) or recall ( $M = .26, SD = .11$ ) tests. Multiple choice test performance was also significantly higher than recall. These results closely align with those in the first set of experiments despite the inclusion of a distractor task immediately before the intervening test (see Figures 4 and 15). Unlike the first set of experiments, no significant interaction was found,  $F(2, 134) = .011, p = .99$ . Performance on yes-no intervening tests was consistently higher than multiple choice performance, indicating

yes-no tests may be a more sensitive measure of memory for previously presented stimuli. The inclusion of a greater number of foils in the multiple choice condition could have led to its greater degree of difficulty. However, three incorrect alternatives per test item is more typical of educational settings than the lower numbers used in other studies comparing the two methodologies. The delay between study list presentation and intervening test led to a slight reduction in intervening test performance from Experiment 1 (see Table 5).



*Figure 15.* Mean Intervening Test Performances by Test Type.



Table 5

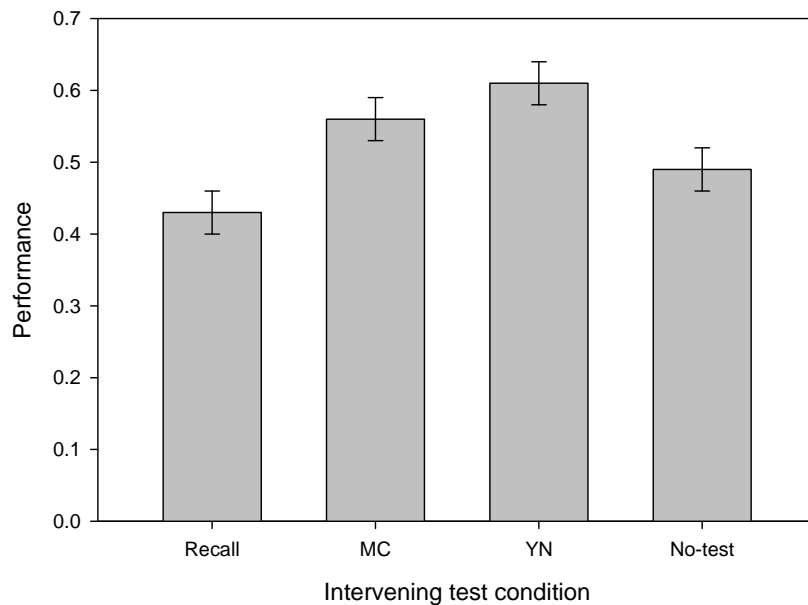
*Mean Correct Identification of Studied Items at Intervening Test in All Experimental Conditions.*

<b>Intervening test</b>	<b>Experiment</b>			
	<b>1 (R/K)</b> <i>Mean (SD)</i>	<b>1 (source)</b> <i>Mean (SD)</i>	<b>2 (R/K)</b> <i>Mean (SD)</i>	<b>2 (source)</b> <i>Mean (SD)</i>
Recall test 1	5.0 (2.5)	5.8 (3.9)	4.4 (2.2)	4.9 (2.9)
Recall test 2	6.0 (1.9)	6.9 (3.0)	6.1 (3.3)	5.4 (2.4)
MC test 1	15.0 (3.9)	17.1 (2.4)	15.0 (2.8)	13.9 (4.0)
MC test 2	13.4 (4.6)	16.7 (3.1)	12.2 (4.4)	12.9 (4.5)
YN test 1	17.2 (2.1)	17.1 (2.4)	16.0 (2.7)	16.2 (2.7)
YN test 2	16.9 (2.8)	16.6 (3.1)	16.0 (3.1)	16.5 (2.8)

*Note. 20 study words per test*

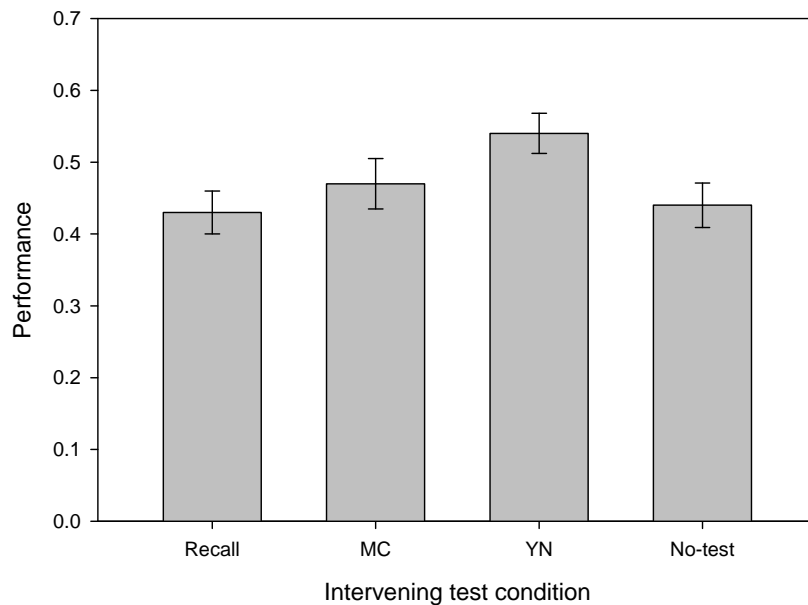
#### *Final process estimation test performance*

The ANOVA for final recognition performance using the *remember-know* task was significant,  $F(3, 99) = 11.94, p < .01$ . Post-hoc comparisons using Bonferroni's correction showed that yes-no condition performance ( $M = .61, SD = .19$ ) was higher than no-test condition ( $M = .49, SD = .19$ ), and recall intervening test condition ( $M = .43, SD = .18$ ) but did not significantly differ from the intervening multiple choice condition ( $M = .56, SD = .19$ ). The difference between the latter two conditions was also significant. Despite the inclusion of a distractor task after study lists, these results closely approximate the results from Experiment 1 (see Figures 5 and 16).



*Figure 16.* Overall Final Estimation Performance by Intervening Test Condition in *Remember-Know* Condition.

An ANOVA of final process estimation test performance using the source memory test was also significant,  $F(3, 102) = 4.60, p < .01$ . The same performance pattern could be seen as in in the *remember-know* condition (see Figures 16 and 17), though only the performance advantage of taking yes-no intervening tests ( $M = .55, SD = .16$ ) over recall ( $M = .43, SD = .18$ ) and no-test ( $M = .44, SD = .19$ ) reached significance. The inclusion of a distractor task following the study list appears to have affected overall performance in the final source memory test, though the pattern of results closely approximates the pattern seen in the *remember-know* final test conditions in Experiments 1 and 2 (see Figures 5, 16, and 17).



*Figure 17. Overall Final Estimation Performance by Intervening Test Condition in Remember-Know condition.*

In both final test conditions, as in Experiment 1, a significant testing effect was found when using yes-no tests. This effect was significant despite both no-test conditions and yes-no conditions involving presentation of the items for a second time. The primary difference appears to be that in the yes-no condition participants were asked to make a decision about whether the item was presented in the initial study trial, while the no-test condition simply passively viewed the words for a second time.

*Recollection and familiarity*

Neither the ANOVA for *remember-know* test condition,  $F(3, 99) = 1.65, p = .18$ , nor the ANOVA for source memory final test condition,  $F(3, 102) = 2.22, p = .09$ , yielded significant results. It seems that the inclusion of a distractor task following the

study lists reduced the near ceiling hit and *remember* rates seen in Experiment 1. However, as in the first set of experiments, the overall comparison of intervening test conditions did not indicate a difference in recollection and familiarity. It does not appear that ceiling effects in hits or *remember* responses led to the failure to find significant differences in recollection between testing conditions in Experiment 1. The pattern of recollection rates across intervening test and no-test conditions was generally convergent with those found in the first experiment (see Table 6). These results indicate that the mere act of taking an intervening test, regardless of test type, does not lead to overall increases in recollection.

Table 6

*Final Recollection Probabilities for Remember-Know and Source Conditions.*

<b>Variable</b>	<b>Condition</b>			
	<b>Recall</b> <i>Probability</i> <i>(SD)</i>	<b>MC</b> <i>Probability</i> <i>(SD)</i>	<b>YN</b> <i>Probability</i> <i>(SD)</i>	<b>No-test</b> <i>Probability</i> <i>(SD)</i>
<i>Remember-know condition</i>				
Raw <i>remember</i>	.49 (.23)	.53 (.25)	.55 (.23)	.55 (.25)
Conditional <i>remember</i>	.61 (.24)	.60 (.26)	.63 (.24)	.65 (.26)
<i>Source memory condition</i>				
Correct source	.38 (.14)	.36 (.15)	.43 (.17)	.39 (.17)
Incorrect Source	.21 (.11)	.23 (.13)	.23 (.11)	.19 (.11)
No source	.23 (.14)	.26 (.16)	.19 (.14)	.25 (.17)

#### IV. GENERAL DISCUSSION

##### *The testing effect in recognition*

The beneficial encoding effects of taking an intervening test, compared to an additional study trial, appear to differ according to type of intervening test. All studies demonstrated an advantage for intervening test conditions when using recognition memory tests as intervening and final tests, despite equating for exposure in the no-test condition. The advantage for taking an intervening yes-no test over the study items compared to additional exposure to the study items reached significance in all conditions of both experiments except the source memory final test condition in Experiment 1.

Interestingly, the source memory final test condition in Experiment 1 is also the only condition that failed to find significantly higher performance on intervening yes-no tests than multiple choice tests. This suggests that performance on intervening tests has an impact on their effectiveness at facilitating later retrieval. This may also explain the consistently lower final recognition performance following intervening recall, as significantly fewer words were identified in intervening recall tests than intervening recognition tests.

Several studies that manipulated the difficulty and number of lures using multiple choice tests have demonstrated that these factors can affect the advantage of multiple choice testing (Roediger & Marsh, 2005, Butler, Marsh, Goode & Roediger, 2007).

However, little is known about how failing to identify an item at intervening test impacts performance on the item during a final recognition test when lures are not reused. An examination of the pattern of misses at final test in the current study supports the importance of intervening test performance. For example, in the remember-know condition in Experiment 1 the probability of missing an item at final test if it was correctly identified at intervening test was approximately .08, however, the probability of missing an item at final test if it was missed at intervening test was .25 (see Table 3).

The importance of intervening test performance is indirectly supported by recent research demonstrating the beneficial effects of feedback. Kang, McDermott, and Roediger (2007) report a significant testing advantage for conditions in which feedback was given following a short-answer intervening test, but not in short answer testing conditions without such feedback. Although the effect of feedback was not significant for intervening multiple choice tests, the authors point out that intervening multiple choice performance was already significantly higher than short answer performance, so fewer opportunities existed for feedback to further impact encoding. Pashler, Cepeda, Wixted, and Rohrer's (2005) results support this contention. Using a paired associates tasks, they found that providing feedback about the correct answer following an incorrect response greatly increased retention, however similar feedback after a correct response did not have any effect on later performance. Wininger (2005) further demonstrated that the beneficial effects of feedback in classroom testing may differ according to the type of feedback given. Thus, the beneficial effects of testing may be contingent on either sufficiently high performance at intervening test, or appropriate feedback following the

test. An examination of the data in the current studies suggests that this may be particularly true when using recognition final tests. These studies suggest that the failure to find significant testing effects in all testing conditions, particularly the recall conditions, may be due to the significantly lower intervening test performance. It is possible that this disadvantage may have been remedied through the use of feedback.

An alternative explanation for the pattern of results obtained in the present experiments involves transfer appropriate processing. According to the transfer appropriate processing view of the testing effect, the testing advantage is the result of the similarity of processes invoked at intervening test and final test. Hence, the encoding effectiveness of taking an intervening test relies on similar mechanisms at work at final test. This explanation suggests that the greater performance in yes-no intervening test conditions is due to the yes-no format being the most similar to final *remember-know* or source memory tasks. In both yes-no and final process estimation tests, items were presented individually at test and participants were asked to make a decision about the item's inclusion on previous study lists. Because recall intervening tests were the least similar to final recognition tests, performance in recall intervening test conditions did not show evidence of a testing effect. This explanation is supported by research on encoding specificity (e.g. Morris, Bransford, & Franks, 1977; see also Tulving and Thompson, 1973), and Duchastel & Nungester's (1982) reported transfer effects in a study using multiple choice and short answer intervening and final tests.

Several recent studies have been cited as support for the view that the benefits of testing occur as a result of more elaborative processing at test (Kang, McDermott, &

Roediger, 2007; Carpenter & DeLosh, 2006), rather than transfer. Confirmatory support for this view could come from studies demonstrating a testing advantage for tests commonly seen as more elaborative (i.e., recall or short answer) regardless of the type of final test. However, as mentioned above, few studies have demonstrated such an effect using recognition final tests. Kang, McDermott, and Roediger's (2007) first experiment demonstrated an advantage for intervening multiple choice tests, regardless of whether the final format was multiple choice or short answer. Additionally, in a direct test of transfer appropriate and elaborative retrieving theories, Carpenter and DeLosh (2006) found no significant effect of test type on final recognition performance. These results indicate that the greater elaboration that is assumed to occur when taking recall or short answer tests do not necessarily lead to greater performance at final test when the final test format is recognition. The results of the present study support this conclusion.

The elaborative processing and transfer appropriate processing views are not mutually exclusive, and both may be important contributors to the testing effect. The theories could be reconciled by suggesting that recognition final tests are simply more susceptible to transfer effects. Thus, the similarity of encoding conditions at intervening and final test play a role in the testing effect, with elaboration at retrieval playing a slightly lessened role when using recognition final tests. However, feedback better modulates elaborative processing in recall intervening tests. This may lead to a greater advantage for more elaborative methods when feedback is given, while also explaining why this advantage is often not found without such feedback (e.g. the current study, Carpenter & Delosh, 2006, Exp. 1; Kang, McDermott, and Roediger, 2007, Exp. 1).



Indeed, several recent studies, including McDaniel, Anderson, Derbish, and Morrisette (2007), and Kang, McDermott, and Roediger's (2007) experiment 2, have demonstrated a significant advantage for recall (short answer) intervening tests when feedback was given.

It is also possible that the short retention interval between intervening and final tests in this study could have reduced the likelihood of finding a significant testing effect in the recall and multiple-choice intervening test conditions. Numerous studies have suggested that longer retention intervals result in more robust testing effects in both recall (e.g. Roediger & Karpicke, 2006b) and recognition (e.g., Wenger, Thompson, & Bartling, 1980) final tests. At short retention intervals, re-study conditions sometimes lead to comparable, or even better, performance, but testing's advantage has been shown to increase with time (for review see Roediger & Karpicke, 2006). Longer delays could also differentially impact recollection and familiarity, though more research is needed to determine whether such an effect could be obtained.

#### *Testing and recollection*

Perhaps most importantly, the results of this study indicate that the notion that taking an intervening test over material leads to improved recollection is overly simplistic. Results obtained using the *remember-know* and source memory process estimation measures indicated no difference in conditional or raw probabilities of recollection between intervening test conditions and no-test conditions. Moreover, further analyses indicated that a significant testing advantage for recollection does exist, but only for items that were correctly recalled at intervening test. Items that were recalled at

intervening test were more likely to be recollected at final test than items that were given additional study time. Although items that were correctly identified in multiple choice or yes-no tests were more likely to be recollected than items that were not correctly identified at intervening test, no consistent recollection advantage existed for these items over items that were given additional study time. It is likely that the greater elaboration in recalling an item resulted in this recollection advantage. This suggests that although a transfer appropriate processing advantage may occur for overall recognition performance, a recollection advantage may occur due to elaborative processing.

The results of the present study demonstrate the effect of taking tests over studied material by comparing intervening test conditions to a condition in which study items were presented a second time for additional exposure. Because neither Chan and McDermott (2007) nor Jones & Roediger (1995) equated for exposure to study items in the no-test conditions, it is possible that much of the increase in recollection in those studies was the result of the additional exposure to study items that occurred at test, rather than the encoding value of the retrieval processes involved in taking the intervening test. The conditional rate of *remember* responses in Chan & McDermott's (2007) intervening recall condition (.66) is comparable to the conditional *remember* rate in Experiments 1 and 2 (Means from .60 to .73). However, Chan & McDermott's no-test conditional *remember* rate in the same study (.57) was lower than any no-test or intervening test condition in the current set of experiments. Additionally, it is likely that the higher intervening recall test performance in Chan & McDermott's (2007) study (Means from .49 to .56) led to their recollection differences between intervening test and no-test

conditions. It is possible that the low intervening recall test performance in these studies (Means from .26 to .32) resulted in too few correctly recalled items at final test to lead to significant recollection differences between testing conditions. The intervening recall performance difference between the current study and Chan and McDermott's study could be due to the four-second word presentation rate during study trials in Chan and McDermott's study, which was twice the amount of time given in the present studies.

These findings further emphasize the importance of intervening test performance in using intervening tests as a tool for improving memory, particularly when using more elaborative intervening tests. Future research is also needed to determine whether this recollection advantage can also be found for items that are not correctly recalled at intervening test if feedback is given, or if such feedback results in recollection improvement for items in intervening recognition test conditions. However, the results of this set of studies clearly demonstrates that any final recollection advantage for taking intervening tests is specific to correctly identified items if no such feedback is given.

A possible limitation of asserting that the recollection advantage for correctly recalled items seen in both Chan & McDermott (2007) and the present study is that items that were correctly recalled at intervening test were likely stronger memories at intervening test than items that were not recalled. As stronger memories they may have been items that were likely to elicit *remember* judgments or correct source judgments at final test even without the benefit of testing. However, this limitation can not be avoided when comparing items that were correctly identified at intervening test with items that were not. A second limitation, particularly of Chan & McDermott's study, is that items

were visible after recall for the remainder of the recall period (90s). Participants could have benefited from the additional exposure to these words during that time. This assumes that participants were sufficiently motivated to take advantage of any additional time to study the items. The present study attempted to control for this potential confound by asking participants to type the recalled words using their keyboards, and limiting the screen space devoted to the recalled words to a small area at the bottom of the screen. Only a few items could be seen at any given time unless participants attempted to scroll up to see the words they had already typed. An informal analysis of all recalled items indicated no recollection advantage for the last few items recalled, those that would have remained visible for the final seconds of the recall interval, compared to those recalled earlier.

Several studies suggest that *remember* and *know* judgments are not actually based on recollection, but on confidence or phenomenological experience (e.g., Wixted & Stretch, 2004; Donaldson, 1996; see also Wixted 2007). The finding that *remember* false alarms are made more quickly than *know* hits has been cited as support for this signal detection interpretation of *remember-know* results (Wixted & Stretch, 2004). This effect was also found in the current study. The finding that incorrect source judgments are also made more quickly than no-source judgments in a similar one-step procedure is also reconcilable from the signal detection position that memory decisions are made along an axis of strength (see Figure 2). Memories for which source information exists, correct or incorrect, should be strong, situated at a fairly high point on the memory strength axis, and accompanied by a high degree of confidence and faster response times. If so,

incorrect source judgments might be made as quickly as correct source judgments. False memories are often made with high confidence, and can occur as quickly as true memories (e.g., Tun, Wingfield, Rosen & Blanchard, 1998; Slotnick & Schacter, 2004). However, more research is needed to better understand the role of confidence in the testing effect, and whether greater confidence following test trials can potentially lead to the increases in *remember* judgments found in these studies. Kang (2008) reported preliminary results indicating that participants in repeated study conditions actually make higher predictions for future recall than participants in repeated test conditions. If so, it is possible that intervening study trials actually lead to greater confidence than intervening tests.

In addition to their convergence in response time patterns, the *remember-know* and source memory procedures were generally convergent in their estimates of recollection and familiarity. Neither procedure demonstrated an effect of intervening test condition, though both indicated a similar pattern of recollection when comparing items that were correctly identified or missed at intervening test with items that were not tested. This may indicate that they are both measuring components of the same underlying construct. Memory for source information is a generally agreed upon component of the recollection process. *Remember* responses, whether they are based on confidence or a conscious experience accompanying recollection, also closely align with this recollection process. Convergence between the *remember-know* procedure and other measures of recollection has been consistently demonstrated, with the exception of response time studies. The present experiments indicate that these response time patterns can also be

found in source memory judgments if using a one-step procedure similar to that used in *remember-know* studies. This may result from participants first determining if recollective detail is present, and responding with *know* or no-source judgments only if it is not.

### *Significance to education*

Perhaps the message most compelling for educational practice is that testing can improve later retention relative to additional study of the material. This is true of both recognition-based tests and more elaborative recall-based tests. However, the impact of testing on later memory is heavily influenced by performance at intervening test. This may be particularly salient when using more elaborative testing methods, such as recall or short answer essay questions. Tests requiring more elaboration also seem to be more sensitive to the benefits of feedback, although this could be due to lower initial test performance in many studies examining this effect. In order to ensure that students benefit from testing, it may be beneficial to either use more elaborative methods and provide feedback, or administer tests with high likelihood for initial student success. Due to the primary function of tests in classroom settings as assessment tools, the former may be a more realistic option.

### Conclusions

Early studies suggested that the testing effect may not occur in recognition (e.g., Hogan & Kitsch, 1971). However, the results of the current studies have demonstrated

that taking intervening tests can improve memory in later recognition tests. These effects likely differ depending on whether an item was correctly identified at intervening test.

Two potential explanations exist for finding a reliable testing effect using yes-no intervening tests. The first is that the results are due to transfer appropriate processing, as the intervening test format most similar to the final recognition test was the yes-no test. Although this study was not designed to compare the transfer appropriate processing and elaborative processing explanations for the testing effect, the significant testing effect using recognition final tests, and advantage for recognition over recall intervening test conditions do not support an elaborative processing view.

Another possible explanation, which has not been as extensively examined using recognition final tests, is that level of intervening test performance modulates the encoding effectiveness of intervening tests. In all conditions except the source memory condition in Experiment 1, performance on yes-no intervening tests was significantly higher than other intervening test formats, and final test performance in the yes-no intervening test conditions was also higher. Recognition final tests may be particularly susceptible to performance differences in intervening tests.

These data also suggest that the simple act of taking intervening tests does not differentially affect recollection and familiarity. Correctly identifying an item on an intervening test leads to greater chance of later recollecting that item at final test compared to items that were not correctly identified at intervening test. However, only correctly recalling an item at intervening recall test leads to a greater recollection advantage at final test compared to re-exposure to study items. In these findings, and the

finding that correct and incorrect recollection-based responding was faster than familiarity-based responding, the two process estimation tests were convergent.



## REFERENCES

- Abbott, E. E. (1909). On the analysis of the factors of recall in the learning process. *Psychological Monographs*, 11, 159–177.
- Anderson, J. R., & Bower, G. H. (1974). A prepositional theory of recognition memory. *Memory & Cognition*, 2, 406-412.
- Atkinson, R. C., Herrmann, D. J., & Wescourt, K. T. (1974). Search processes in recognition memory. In R.L. Solso (Ed.) *Theories in Cognitive Psychology: The Loyola Symposium*. New York, John Wiley & Sons.
- Atkinson, R. C. & Juola, J. F. (1973). Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.) *Attention and Performance*, Vol 4, New York, Academic Press.
- Atkinson, R. C., & Juola, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R.C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology: Vol. I. Learning, memory & thinking*. San Francisco: Freeman.
- Atkinson, R. C. & Wescourt, K. T. (1975). Some remarks on a theory of memory. In P. M. A. Rabbitt & S. Dornic (Eds.) *Attention and Performance*, Vol 5, London, Academic Press.

- Bahrick, H. P. (1970). Two-phase model for prompted recall. *Psychological Review*, 77, 215-222.
- Blackwell, D. (1953). Equivalent comparisons of experiments. *Annals of Mathematics and Statistics*, 24, 265-272.
- Bower, G. H., Clark, M. C., Lesgold, A. M., & Winzenz, D. (1969). Hierarchical retrieval schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior*, 8, 323-343.
- Butler, A. C., Marsh, E. J., Goode, M. K., & Roediger, H. L., III. (2006). When additional multiple-choice lures aid versus hinder later memory. *Applied Cognitive Psychology*, 20, 941-956.
- Butler, A. C., & Roediger, H. L., III. (2007). Testing improves longterm retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, 19, 514-527.
- Calkins, M. W. (1894). Association. *Psychological Review*, 1, 476-483.
- Cameron, L. (2002). Measuring vocabulary size in English as an additional language. *Language Teaching Research*, 6, 145-173.
- Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition*, 34, 268-276.
- Cooper, A. J. R., & Monk, A. (1976). Learning for recall and learning for recognition. In J. Brown (Ed.), *Recall and recognition* (pp. 131-156). New York: Wiley.

- Curran, T., & Hintzman, D. L. (1995). Violations of the independence assumption in process dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 531–547.
- Dale, H.C.A. (1967). Response availability and short term memory. *Journal of Verbal Learning and Verbal Behavior*, *6*, 47-48.
- 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.
- Deffenbacher, K. A., Leu, J. R., & Brown, E. L. (1981). Memory for faces: Testing method, encoding strategy, and confidence. *American Journal of Psychology*, *94*, 13–26.
- Dewhurst, S. A., & Conway, M. A. (1994). Pictures, images, and recollective experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1088–1098.
- Dewhurst, S. A., Holmes, S.J., Brandt, K. R., & Dean, G. M. (2006). Measuring the speed of the conscious components of recognition memory: Remembering is faster than knowing. *Consciousness and Cognition*, *15*, 147-152.
- Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory and Cognition*, *24*, 523-533.
- Donaldson, W., MacKenzie, T. M., & Underhill, C. F. (1996). A comparison of recollective memory and source monitoring. *Psychonomic Bulletin & Review*, *3*, 486–490.

- Duchastel, P. C. (1981). Retention of prose following testing with different types of test. *Contemporary Educational Psychology, 6*, 217–226.
- Duchastel, P. C., & Nungester, R. J. (1982). Testing effects measured with alternate test forms. *Journal of Educational Research, 75*, 309–313.
- Dunn, J. (2004). Remember-know: A matter of confidence. *Psychological Review, 111*, 524-542.
- Eagle, M., & Leiter, E. (1964). Recall and recognition in intentional and incidental learning. *Journal of Experimental Psychology, 68*, 58-63.
- Ebbinghaus, H. (1885). *Memory: A contribution to experimental psychology*. Translated by Ruger, H. A., & Bussenius, C.E.. New York: Dover.
- Eichenbaum, H., Yonelinas, A. R., Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience, 30*, 123-152.
- Feenan, K., & Snodgrass, J. (1990). The effect of context on discrimination and bias in recognition memory for pictures and words. *Memory & Cognition, 18*, 515-527.
- Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition, 16*, 309–313.
- Gardiner, J. M., & Java, R. I. (1990). Recollective experience in word and nonword recognition. *Memory & Cognition, 18*, 23–30.
- Gates, A. I. (1917). Recitation as a factor in memorizing. *Archives of Psychology, 6*, 40.
- Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91*, 1-65.

- Glanzer, M., Kim, K., Hilford, A., & Adams, J. K. (1999). Slope of the receiver-operating characteristic in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 500–513.
- Glover, J. A. (1989). The “testing” phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, *81*, 392–399.
- Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural contexts: On land and underwater. *British Journal of Psychology*, *66*, 325–331.
- Green, D. M., & Moses, F. L. (1966). On the equivalence of two recognition measures of short-term memory. *Psychological Bulletin*, *66*, 228–234.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Gronlund, S. D., & Ratcliff, R. (1989). Time course of item and associative information: Implications for global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 846–858.
- 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.

- Healy, M. R., Light, L. L., Chung, C. (2005). Dual-process models of associative recognition in young and older adults: Evidence from receiver operating characteristics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 768-788.
- Hicks, J. L., Marsh, R. L., & Ritschel, L. (2002). The role of recollection and partial information in source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 503-508.
- Higham, P.A. & Vokey, J.R. (2004). Illusory recollection and dual-process models of recognition memory. *The Quarterly Journal of Experimental Psychology*, 57, 714-744.
- Hintzman, D. L., & Caulton, D. A. (1997). Recognition memory and modality judgments: A comparison of retrieval dynamics. *Journal of Memory and Language*, 37, 1–23.
- Hirshman E., Master, S. (1997). Modeling the conscious correlates of recognition memory: Reflections on the remember-know paradigm. *Memory and Cognition*, 25, 345-351.
- Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior*, 10, 562–567.
- Jacoby, L. L. (1984). Incidental versus intentional retrieval: Remembering and awareness as separate issues. In L. R. Squire & N. Butters (Eds.), *Neuropsychology of memory* (pp. 145–156). New York: Guilford Press.

- 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. (1998). Invariance in automatic influences of memory: Toward a user's guide for the process-dissociation procedure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 3–26.
- Jacoby, L. J., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, 110, 306-340.
- Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General*, 118, 126–135.
- Jacoby, L. J., Yonelinas, A. P., & Jennings, J. M. (1997). The relation between conscious and unconscious (automatic) influences: A declaration of independence. In E. Jonathan, D. Cohen, & W. Schooler (Eds.), *Scientific approaches to consciousness* (pp. 13–47). Hillsdale, NJ: Erlbaum.
- James, W. (1890). *Principles of Psychology*. Cambridge, MA: Harvard University Press.
- Java, R. I., Gregg, V. H., & Gardiner, J. M. (1997). What do people actually remember (and know) in “remember/know” experiments? *European Journal of Cognitive Psychology*, 9, 187–197.
- Jones, T.C. (2006). Editing (out) generated study words in a recognition exclusion task. *Memory*, 14, 712-729.

- Jones, T. C., & Jacoby, L. L. (2001). Feature and conjunction errors in recognition memory: Evidence for dual-process theory. *Journal of Memory and Language*, 45, 82-102.
- Jones, T. C., & Roediger, H. L. (1995). The experiential basis of serial position effects. *European Journal of Cognitive Psychology*, 7, 65-80.
- Joordens, S., & Merickle, P.M. (1993). Independence or redundancy? Two models of conscious and unconscious influences. *Journal of Experimental Psychology: General*, 122, 462-467.
- Juola, J.F., Fischler, C.T., Wood, C.T., & Atkinson, R.C. (1971). Recognition time for information in long-term memory. *Perception and Psychophysics*, 10, 8-14.
- Kang, S. (2008). *Improving learning and memory through testing*. Paper presented at a meeting at Auburn University, Auburn, AL.
- Kang, S. H. K., McDermott, K. B., & Roediger, H. L., III. (2007). Test format and corrective feedback modulate the effect of testing on memory retention. *European Journal of Cognitive Psychology*, 19, 528-558.
- Kelley, C. M., & Jacoby, L. L. (2000). Recollection and familiarity: Process-dissociation. In E. E. Tulving, E. Ferguson, & I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 215–228). New York, NY: Oxford Univ.Press.
- Kelley, R., & Wixted, J. T. (2001). On the nature of associative information in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 701–722.



- Kintsch, W. (1967). Memory and decision aspects of recognition learning. *Psychological Review*, 74, 496-504.
- Kintsch, W. (1968). An experimental analysis of single stimulus tests and multiple-choice tests of recognition memory. *Journal of Experimental Psychology*, 76, 1-6.
- Kintsch, W. (1970). Models for free recall and recognition. In Norman, D. (Ed.) *Models of Human Memory*. New York: Academic Press.
- Kojic-Sabo, I., Lightbown, P. (1999). Students' approaches to vocabulary knowledge and their relationship to success. *The Modern Language Journal*, 83, 176-192.
- Kroll, N. E., Yonelinas, A. P., Dobbins, I. G., & Frederick, C. M. (2002). Separating sensitivity from response bias: Implications of comparisons of yes-no and forced-choice tests for models and measures of recognition memory. *Journal of Experimental Psychology; General*, 131, 240-254.
- Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Lachman, R., & Tuttle, A. V. (1965). Approximations to English and short-term memory: Construction or storage? *Journal of Experimental Psychology*, 70, 386-393.
- LeCompte, D. C. (1995). Recollective experience in the revelation effect: Separating the contributions of recollection and familiarity. *Memory & Cognition*, 23, 324-334.
- Light, L. L., & Prull, M. (1995). Aging, divided attention, and repetition priming. *Swiss Journal of Psychology*, 54, 87-101.
- Lockhart, R. S. (1975). The facilitation of recognition by recall. *Journal of Verbal Learning and Verbal Behavior*, 14, 253-258.

- Lockhart, R. S. (2000). Methods of memory research. In E. Tulving & F.I.M. Craik's (Eds.) *The Oxford Handbook of Memory*. Oxford: Oxford University Press.
- Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York: NY.
- Mandler, G. (1972). Organization and Recognition. In E. Tulving & W. Donaldson (Eds.) *Organization of Memory*, New York, Academic Press.
- Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, 87, 252-271.
- Mandler, G., & Boeck, W. (1974). Retrieval processes in recognition. *Memory & Cognition*, 2, 613-615.
- Mandler, G., Pearlstone, Z., & Koopmans, H. S. (1969). Effects of organization and semantic similarity on recall and recognition. *Journal of Verbal Learning and Verbal Behavior*, 8, 410-423.
- Mandler, G., & Rabinowitz, J. C. (1981). Appearance and reality: Does a recognition test really improve subsequent recall and recognition? *Journal of Experimental Psychology: Human Learning and Memory*, 7, 79-90.
- McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, 19, 494-513.
- McDaniel, M.A., Kowitz, M.D., & Dunay, P. K. (1989). Altering memory through recall: The effects of cue-guided retrieval processing. *Memory & Cognition*, 17, 423-434.

- McDaniel, M. A., & Masson, M. E. J. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 371–385.
- McDougall, R. (1904). Recognition and Recall. *Journal of Philosophical Psychology and Scientific Methods*, *1*, 229-233.
- McElree, B., Dolan, P. O., & Jacoby, L. L. (1999). Isolating the contributions of familiarity and source information to item recognition: A time course analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 563–582.
- Meiser, T., & Sattler, C. (2007). Boundries of the relation between conscious recollection and source memory for perceptual details. *Consciousness and Cognition*, *16*, 189-210.
- Mitchell, K. J., Johnson, M. K. (2000). Source monitoring: Attributing mental experinences. In E. Tulving & F. I. M. Craik (Eds) *The Oxford Handbook of Memory*. Oxford: Oxford University Press.
- Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer-appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 519–533.
- Murdock, B.B. (1968). Modality effects in short-term memory; Storage or retrieval? *Journal of Experimental Psychology*, *77*, 79-86.
- Parks, C. M., & Yonelinas, A. P. (2007). Moving beyond pure signal-detection models: Comment on Wixted (2007). *Psychological Review*, *114*, 188-202.

- Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 3-8.
- Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory and Cognition*, *21*, 89–102.
- Ranganath, C., Yonelinas, A. P., Cohena, M. X., Dy, C. J., Tom, S. M., D'Esposito, M. (2003). Dissociable correlates of recollection and familiarity within the medial temporal lobes. *Neuropsychologia*, *42*, 2-13.
- Richardson, J. T. E. (1985). The effects of retention tests upon human learning and memory: A historical review and an experimental analysis. *Educational Psychology*, *5*, 85–114.
- 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., III, & Karpicke, J. D. (2006). Test enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255.
- Roediger, H. L., III, & Marsh, E. J. (2005). The positive and negative consequence of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1155–1159.
- Roediger, H. L., III, & McDermott, K. B. (1993). Implicit memory in normal human subjects. *Handbook of Neuropsychology*, *8*, 63–131.

- Rotello, C. M., Macmillan, N. A., Hicks, J. L., & Hautus, M. J. (in press). Interpreting the effects of response bias on remember-know judgments using signal-detection and threshold models. *Memory and Cognition*.
- Rotello, C. M., MacMillan, N. A., Reeder, J. A. (2004). Sum-difference theory of remembering and knowing; A two-dimensional signal detection model. *Psychological Review*, *111*, 588-616.
- Rotello, C. M., Macmillan, N. A., Reeder, J. A., & Wong, M. (2005). The remember response: Subject to bias, graded, and not a process-pure indicator of recollection. *Psychonomic Bulletin & Review*, *12*, 865–873.
- Rotello, C. M., MacMillan, N. A., & Van Tassel, G. (2000). Recall-to-reject in recognition: Evidence from ROC curves. *Journal of Memory and Language*, *43*, 67-88.
- Slotnick, S. D., & Dodson, C. S. (2005). Support for a continuous (single-process) model of recognition memory and source memory. *Memory and Cognition*, *33*, 151-170.
- Slotnick, S.D., & Schacter, D.L. (2004). A sensory signature that distinguishes true from false memories. *Nature Neuroscience*, *7*, 664-672.
- Smith, D. L. (2006). *Development and validation of the Auburn Psychology Term Test (APTT)*. Unpublished masters thesis, Auburn University.
- Smith, D. L., & Barker, L. (under review). Using yes-no recognition tests to assess student memory for course content.
- Smith, S. M. Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory and Cognition*, *6*, 342-353.

- Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology*, 30, 641–656.
- Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 1379-1396.
- Strong, E. K. (1912). The effect of length of series upon recognition memory. *Psychological Review*, 19, 447-462.
- Toth, J. P. (1996). Conceptual automaticity in recognition memory: Levels-of-processing effects on familiarity. *Canadian Journal of Experimental Psychology*, 50, 123–138.
- Tulving, E. (1972). Episodic and semantic memory. In E. Tulving and W. Donaldson (Eds.) *Organization of Memory*, 381-403. New York: Academic Press.
- Tulving, E. (1976) Euphoric processes in recall and recognition. In J. Brown (Ed.) *Recall and Recognition*. London: John Wiley & Sons.
- Tulving, E. (1983). *Elements of episodic memory*. New York: Oxford University Press.
- Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, 26, 1-12.
- Tulving, E., & Thompson, D. M. (1971). Retrieval processes in recognition memory: Effects of associative context. *Journal of Experimental Psychology*, 87, 116-124.
- Tulving, E., Thompson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80, 352-373.

- Tun, P. A., Wingfield, A., Rosen, M. J., & Blanchard, L. (1998). Response latencies for false memories: gist-based processes in normal aging. *Psychology and Aging, 13*, 230-241.
- Watkins, M. J., & Tulving, E. (1975). Episodic memory: When recognition fails. *Journal of Experimental Psychology: General, 104*, 5-29.
- Wagner, A. D., Stebbins, G. T., Masciari, F., Fleischman, D. A., & Gabrieli, J.D.E. (1998). Neuropsychological dissociation between recognition familiarity and perceptual priming in visual long-term memory. *Cortex, 34*, 493-511.
- Wenger, S. K., Thompson, C. P., & Bartling, C. A. (1980). Recall facilitates subsequent recognition. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 135-144.
- Westerman, D. L. (2001). The role of familiarity in item recognition, associative recognition, and plurality recognition on self-paced and speeded tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 723-732.
- Wickelgren, W. A., & Norman, D. A. (1966). Strength models and serial position in short-term recognition memory. *Journal of Mathematical Psychology, 3*, 316-347.
- Williams, E. J. (1949). Experimental designs balanced for the estimation of residual effects of treatments. *Australian Journal of Scientific Research, 2*, 149-168.
- Wininger, S. R. (2005). Using your tests to teach: Formative summative assessment. *Teaching of Psychology, 32*, 164-166.
- Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review, 114*, 152-176.

- Wixted, J. T., & Stretch, V. (2004). In defense of the signal detection interpretation of remember/know judgments. *Psychonomic Bulletin and Review*, *11*, 616-641.
- Yntema, D. B., & Trask, F. P. (1963). Recall as a search process. *Journal of Verbal Learning and Verbal Behavior*, *2*, 65-74.
- Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1341-1354.
- Yonelinas, A. P. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition*, *25*, 747-763.
- Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recognition and source-memory judgments: A formal dual-process model and an analysis of receiver operating characteristics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1415-1434.
- Yonelinas, A. P. (2001). Consciousness, control, and confidence: The 3 Cs of recognition memory. *Journal of Experimental Psychology: General*, *130*, 361-379.
- 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., Hockley, W. E., Murdock, B. B. (1992). Tests of the list strength effect in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 345-355.



Yonelinas, A. P., Kroll, N. E. A., Dobbins, I., Lazzara, M., & Knight, R. T. (1998).

Recollection and familiarity deficits in amnesia: Convergence of remember-know, process dissociation, and receiver operating characteristic data. *Neuropsychology*, *12*, 323–339.

Yonelinas, A. P., Otten, L. J., Shaw, K. N., & Rugg, M. D. (2005). Separating the brain regions involved in recollection and familiarity in recognition memory. *The Journal of Neuroscience*, *25*, 3002-3008.

## APPENDICES

## Appendix A

### *Footnotes*

*I*; Signal detection theory predicts that ROC curves will be curvilinear due to the graded nature of recognition memory. This prediction has held true in virtually all recognition memory experiments conducted using this procedure. In part due to these accurate ROC predictions, signal detection theory has held a prominent position in recognition memory literature and the idea of memory thresholds were largely abandoned. Threshold theories predict linear ROC curves and U-shaped z-ROCs due to their assumption that an item either exceeds a recognition threshold or does not, and for over thirty years no linear ROC curves were found in recognition memory studies. Yonelinas and colleagues (e.g., Yonelinas, 1994; Yonelinas, Kroll, Dobbins, Lazzara, & Knight, 1998; Yonelinas, 2002; Parks & Yonelinas, 2007) were first to propose a dual process account of recognition memory that accurately described ROC data. According to this theory, now referred to as the Dual-Process Signal Detection Theory (DPSDT), familiarity strength exists on a continuum, and is well defined by an equal variance signal detection framework. Recollection, however, operates in manner consistent with a threshold framework. An item is said either to be recollected, thus exceeding the threshold, or not recollected. According to this model, the continuous nature of familiarity leads to curvilinear ROCs, while the threshold properties of recollection lead to their asymmetry. This theory appears to describe ROC data as effectively as the unequal variance signal detection theory (but see Wixted 2007). If this threshold assertion is correct, ROC curves constructed in situations in which a greater number of items are recollected should

become more linear and z-ROCs should become less linear and more U-shaped. Studies of associative and source memory provide such a test of the theory's predictions, as items for which associative or source information is retrieved should rely more heavily on recollection. Yonelinas (1997) was the first to demonstrate such linear ROC curves and U-shaped z-ROCs in associative recognition, in which participants were asked to supply a rating of their confidence that the two presented words were presented together at study. Yonelinas (1999) demonstrated similar findings in another series of source memory experiments in which ROCs were generated by asking participants to rate their confidence in whether a presented item was from one list or presentation modality or another. Rotello, Macmillan, and Van Tassel (2000) also found linear ROCs in a plurality reversed study. Since these initial studies, numerous replications of linear ROCs and U-shaped z-ROCs have been conducted (e.g. Healey, Light, & Chung, 2005; Slotnick & Dodson, 2005). Such findings provided an enormous amount of support for the assertions of the dual process signal detection theory. Particularly since signal detection theories assume ROCs will remain curvilinear due to the continuous nature of memory strength, and source or associative recognition should not be an exception. DPSDT proponents tend to focus on the confirmatory data of U-shaped z-ROCs in associative and source recognition studies, while UVSDT proponents tend to focus on the curvilinearity of nearly all ROC data.

2 and 3; Articles by Witasek (1907) and Katzaroff (1908) are presented as secondary sources because they are unavailable in English, likely a result of their original publication in dated foreign sources.

4; Response times below 500ms were over two standard deviations below the overall mean, and were deemed too fast for the participant to have made adequate decisions in a 1-step source or *remember-know* procedure. Participants who responded this quickly more than once per final process estimation test were not included in the analyses.

5; An examination of items that were recalled at intervening test revealed no significant relation between number of participants who recalled an item and any measurable characteristic, including number of syllables, letters, Kucera-Francis frequency, concreteness, or imagery.

## Appendix B

### *Word lists*

a	b	c	d	e	F	g
QUARTER	FILE	CHAIN	TIRE	GIANT	CAREER	GRIP
CUP	GLORY	MOON	ARRIVAL	PAIR	MOOD	TOOL
LIBRARY	CHIN	SIGNAL	OBJECT	BONE	ARTICLE	CORE
RENT	SHAPE	FORT	TRIM	TRUCK	SELL	DIGNITY
JOIN	EMOTION	BAY	CIRCLE	JURY	VICE	TRAVEL
PERMIT	FISH	MOLD	THEME	AUTHOR	PAGE	COVER
WORKER	DESPAIR	SPOKE	PASS	LUCK	FREIGHT	WIND
PARADE	SALARY	OPENING	GUARD	TITLE	FOOL	HERO
TREND	SKY	HEIGHT	STRAIN	SLIDE	BEAT	CELLAR
HURT	SUIT	LOSS	LIFT	WAGE	ROUGH	SHOOT
CAPTAIN	HIGHWAY	DISPUTE	SCHEME	CALM	OUTCOME	SHOP
HIDE	LOYALTY	ADULT	FASHION	REAR	COMMAND	ARC
NARROW	PUPIL	WINTER	WEAR	GAIN	EVENT	LEAN
INCH	POUND	CIRCUIT	PAYMENT	TILL	ATOM	CHARM
GUIDE	CLAIM	PLUG	WELCOME	ROUND	DAWN	IMPACT
BRIEF	SESSION	PRIZE	FLOWER	BEAM	PALM	RING
TASTE	WEATHER	MYSTERY	DREAM	GRADE	HATE	SAUCE
CRISIS	POET	DRESS	ROOF	MIXTURE	TRACTOR	YARD
ESCAPE	IRON	HEAT	GRAY	THROW	POCKET	SONG
CURVE	OPINION	FELL	MILE	MALE	CORN	ORDERLY

<i>h</i>	<i>i</i>	<i>j</i>	<i>k</i>	<i>l</i>	<i>N</i>	<i>m</i>
QUIET	WATCH	DRAW	SWEET	TREAT	CHEST	CAPITAL
SOFT	ESTATE	NOVEL	DOLLAR	GALLERY	ANIMAL	IMPULSE
DANCE	COLONEL	PIONEER	OXYGEN	JUDGE	CONTENT	CLOUD
EXTRA	PHYSICS	STERN	MINOR	PATENT	PAINT	GESTURE
STORE	PRAIRIE	FINISH	PORCH	SPITE	CONCERT	BLOCK
MOTOR	MYTH	PALE	PICK	OCEAN	COAL	STAKE
LOCK	PLATE	ASPECT	WAIT	RULE	COLUMN	CROWD
PRIMARY	COUNT	WIN	WASTE	CLEAN	POVERTY	STYLE
SHEAR	SYMBOL	ROUTE	CULTURE	RITUAL	AIM	MESSAGE
LAWYER	FOAM	SOLDIER	CARD	WEALTH	DOCTOR	LOAN
SISTER	NOTICE	SAVAGE	HARMONY	DUTY	IDEAL	VACUUM
REPAIR	TALE	WISE	AUTUMN	POWDER	ROCK	GATE
SEAT	VERSE	TRIP	BOATING	BEAR	COMEDY	TALENT
METAL	CHINA	WEIGHT	MAIL	EMPIRE	BRANCH	CITIZEN
SEARCH	ANGLE	MERIT	SUSPECT	BRUSH	GOLF	LOOP
SMELL	PLOT	QUEEN	QUARREL	STABLE	SIGN	TRIBUTE
FRAME	FIGHT	FUN	COOL	SPLIT	REMARK	JOURNAL
GULF	CARBON	COAST	EXPRESS	POST	SAFE	COOK
LUMBER	STORM	BAR	ROLL	QUICK	SELF	DECK
JUNIOR	TREATY	COMFORT	MATCH	RAIN	WALK	CATTLE

<i><b>o</b></i>	<i><b>p</b></i>	<i><b>q</b></i>	<i><b>r</b></i>	<i><b>s</b></i>	<i><b>t</b></i>	<i><b>u</b></i>
TOUCH	LOBBY	CRY	FENCE	TRACK	CURE	TERM
WITNESS	FAINT	LADY	LANE	HARM	LOAD	MANKIND
SUCCESS	PORT	WAVE	PLAIN	OIL	SHORE	SALE
FOIL	GUEST	MEASURE	DRINK	PORTION	MAGIC	ARTIST
CHART	STEEL	DAMAGE	JET	WIDOW	BOWL	FLESH
DEBATE	KNEE	TRAIN	MATE	PRODUCT	FOG	HOLE
PRIME	SAVE	EMPTY	BALANCE	PULL	SWIFT	SMILE
TRIUMPH	NEST	MOTIVE	BORDER	NAVY	PAINTER	DISEASE
SOAP	LUNCH	BID	MOVIE	CAPE	SHOCK	STEM
EASE	SUITE	SPEAKER	SIGHT	KINGDOM	PALACE	SKILL
SPHERE	COUSIN	PROTEST	DEVICE	NATIVE	VICTORY	EXTREME
BATTLE	FACTORY	THICK	MUSCLE	GIFT	NECK	RANCH
CASH	BENEFIT	KING	APPEAL	SWEAT	INSIGHT	ADVICE
ALERT	WISDOM	DRAMA	LAUGH	PATIENT	OWNER	SHELL
PHRASE	CONCEPT	GUY	DELIGHT	PASSAGE	SHAME	ENGINE
AVENUE	NET	SEA	DEAR	SHADOW	LIQUID	CAFE
FILL	LUXURY	DUST	BET	EDITION	LESSON	BOX
CRAFT	CONCERN	VILLAGE	PRODUCE	TEAM	AGENCY	LEADER
TRUST	WHEEL	ORIGIN	BOTHER	COLONY	CREW	DESK
JUMP	WASH	DRILL	YOUTH	MASTER	BARREL	BREATH



<b>v</b>	<b>w</b>	<b>x</b>
BULLET	SINK	ZERO
DOZEN	AUNT	BLONDE
WILD	BANK	CROSS
SUPPER	BUDGET	MOTION
CELL	UNIFORM	RISK
FELLOW	ROOT	FINANCE
TAPE	CABIN	CONTACT
JOY	RELIEF	SALT
VEIN	TIP	PILE
CLOTH	CHEEK	MUSTARD
SCALE	SOIL	THROAT
VALLEY	BASE	MINE
BOND	EXCUSE	SITE
GOAL	PHASE	SPEECH
BRAIN	TUBE	BUREAU
INJURY	GENIUS	FOOT
BREAK	TRACE	GUILT
TRAIL	EDGE	COURAGE
TASK	CAPITOL	REVENUE
VIRTUE	MERCY	AMATEUR

## Appendix C

### *Word list means*

<i>LIST</i>	<i>Concreteness</i>	<i>Imagery</i>	<i>Kucera-Francis</i>		
			<i>Frequency</i>	<i>Letters</i>	<i>Syllables</i>
1a	434.95	464.75	50.6	5.25	1.5
2b	449.55	490.2	52.4	5.35	1.8
3c	477.6	509.7	56.6	5.1	1.4
4d	424.63	468.73	48.89	5.31	1.42
5e	440.1	479.75	46.6	4.6	1.25
6f	440.35	461.85	45.3	5.05	1.45
7g	473.5	509.7	42.9	4.75	1.45
8h	474.55	484.15	54.4	5.2	1.55
9i	470	483.6	42.4	5.35	1.5
10j	418.4	478.45	48.05	4.95	1.45
11k	447.35	479.2	44.15	5.45	1.6
12l	452.35	484.65	47.15	5.15	1.5
13m	457.9	506	53.75	5.15	1.6
14n	463.1	483.95	44.65	5.6	1.6
15o	431.35	467.95	41	5.25	1.4
16p	481.75	489.75	42.4	5.2	1.55
17q	459.52	494.36	56.57	5.05	1.57
18r	446.05	477.6	47.05	5.2	1.45
19s	481.95	499.55	47.9	5.4	1.7
20t	455.95	498.9	37.6	5.2	1.6
21u	485.7	505.35	50.25	5.15	1.4
22v	474.35	512.8	48.2	4.85	1.4
23w	467.45	489.1	42	4.9	1.5
24x	443.35	476.35	43.3	5.5	1.55
Overall M	456.3232	487.3502	47.25515	5.165351	1.508333
Overall SD	19.10863	14.764	5.02689	0.236488	0.11436

## Appendix D

### *Sample instructions*

Thank you for participating in this study. Because this study is important to us, we ask that you make sure that you have chosen a time in which you will be able to participate free of distractions. Please turn your cell-phones off, and refrain from talking or doing other work during the next hour and fifteen minutes. If you are unable to follow these instructions you will be asked to leave immediately.

During this study you will be presented with lists of words, and your memory for these words will be tested using various memory tests. This study consists of 4 blocks. In each block you will be presented with two separate lists of words and will be tested over these words. You will also be asked to generate and remember numbers and perform basic math problems. You will use your keyboard to respond to all tasks. At the end of each block you will receive a final test over the words in that block. In this final test you will be asked to:

#### **RK**

*Identify whether you remember specific details of seeing the item on the list, the word simply seems familiar, so you know it was presented before, or the word was not presented on the previous lists and is new.*

***If you can remember specific details of seeing the item on the list, and could report what these details are if asked, you will press the 4 key.***

***If you do not remember specific details of seeing the word, but know it was presented in a list, press the 5 key.***

***If you do not believe the item was on a study list, press the 6 key.***

#### **SM**

*Identify whether the item was presented before, and if so, whether you can remember which list it was on.*

***If the item was on the first list of that block you will press the 1 key.***

***If the item was on the second list of that block you will press the 2 key.***

***If the item was on a previous list, but you don't remember which list, press the 3 key.***

***If the item was not presented on the previous lists, press the 4 key.***

At the end of each block you will receive instructions that you have finished the block, and if you would like to take a short break, and can do so without distracting others, you are welcome to. Once you have finished a block, you will not be tested over any of the items from that block again.

Instructions will be presented on your computer screen each step of the way. Please pay close attention to these instructions, making sure that you understand them before moving on.

Do you have any questions? Do these instructions make sense to you? (I did not proceed until everyone signaled that they did)

Please take the amount of time you need to respond with your best answer, and no more. You may now begin. The first set of instructions is on your screen.

## Appendix E

### *Sample screen stills*

In this block, you will be asked to view two lists of words. You will later be tested over your memory for these words.

Several types of tests and different tasks will follow the lists of words. You will be given instructions on how to complete each of these tasks.

Please pay close attention to the words, and answer the test questions as accurately as you can.

Press the <space bar> to continue

recall

Which of the following words appeared on the last list?

1. sweet
2. dollar
3. bay
4. oxygen